Contextual modulation of visual variability: perceptual biases over time and across the visual field

Marta Suárez Pinilla

Submitted for the degree of Doctor of Philosophy

University of Sussex - September 2018

Statement

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Marta Suárez-Pinilla

Summary

CONTEXTUAL MODULATION OF VISUAL VARIABILITY: PERCEPTUAL BIASES OVER TIME AND ACROSS THE VISUAL FIELD

The visual system extracts statistical information about the environment to manage noise, ensure perceptual stability and predict future events. These summary representations are able to inform sensory information received in subsequent times or in other regions of the visual field. This has been conceptualized in terms of Bayesian inference within the predictive coding framework. Nevertheless, contextual influence can also drive anti-Bayesian biases, as in sensory adaptation.

Variance is a crucial statistical descriptor, yet relatively overlooked in ensemble vision research. We assessed the mechanisms whereby visual variability exerts and is subject to contextual modulation over time and across the visual field.

Perceptual biases over time: serial dependence (SD)

In a series of visual experiments, we examined SD on visual variance: the influence of the variance of previously presented ensembles in current variance judgments. We encountered two history-dependent biases: a positive bias exerted by recent presentations and a negative bias driven by less recent context. Contrary to claims that positive SD has low-level sensory origin, our experiments demonstrated a decisional bias requiring perceptual awareness and subject to time and capacity limitations. The negative bias was likely of sensory origin (adaptation).

A two-layer model combining population codes and Bayesian Kalman filters replicated positive and negative effects in their approximate timescales.

Perceptual biases across the visual field: Uniformity Illusion (UI)

In UI, presentation of a pattern with uniform foveal components and more variable peripheral elements results in the latter taking the appearance of the foveal input. We studied the mechanistic basis of UI on orientation and determined that it arose without changes in sensory encoding at the primary visual cortex.

Conclusions

We studied perceptual biases on visual variability across space and time and found a combination of sensory negative and positive decisional biases, likely to handle the balance between change sensitivity and perceptual stability.

Collaborations and Publications arising from this Thesis

A large part of the experimental work reported in this thesis has been published or accepted for publication in scientific journals. Specifically, this thesis incorporates extensive content (including analyses, text, tables and figures) from the following articles:

1. Marta Suárez-Pinilla, Anil K. Seth, Warrick Roseboom; Serial dependence in the perception of visual variance. *Journal of Vision* 2018;18(7):4. doi: 10.1167/18.7.4.

2. Marta Suárez-Pinilla, Anil K. Seth, Warrick Roseboom; The illusion of uniformity does not depend on the primary visual cortex: evidence from sensory adaptation. *i-Perception*, 9(5), 1–12. doi: 10.1177/2041669518800728.

Part II: Contextual Modulation Over Time: Serial Dependence in the Perceptual Processing of Visual Variance incorporates content from the first article ('Serial dependence in ...'), specifically concerning the Introduction (Part II Chapter 1), Experiments 1, 2A, 2B and 3, and the majority of the Discussion (Part II Chapter 6).

Part III: Contextual Modulation Across the Visual Field: Sensory Processing Under the Uniformity Illusion (UI) incorporates content from the second article ('The illusion of uniformity ...'), specifically concerning Experiment 5 (Part III Chapter 7) and the majority of the Discussion (Part III Chapter 9).

My contribution as first author of both articles included designing research, performing research, analysing the data and writing the manuscripts. All text, figures and research are my own. The two co-authors (main and second supervisor of this thesis work) contributed by designing and advising research and revising the manuscripts. The random dot kinematograms employed as visual stimuli in the serial dependence study

(reported in Part II) were programmed on the basis of previous Matlab scripts designed by Dr. Acer Chang.

Table of Contents

List of Tables

List of Figures

List of abbreviations

2AFC	two-alternative forced choice
2IFC	two-interval forced choice
BF	Bayes factor
cd/m^2	candela(s) per square metre
CFS	continuous flash suppression
cm	centimetre(s)
CW	clockwise
dva	degree(s) of visual angle
DoG	derivative of Gaussian
GCW	global clockwise
GXCW	global counter-clockwise
HR	heart rate
Hz	hertz(s)

ISI	inter-stimulus interval
LCW	local clockwise
LMM	linear mixed-effects model
LXCW	local counter-clockwise
max	maximum
ms	millisecond(s)
NP	neural population
PAS	Perceptual Awareness Scale
PDF	probability density function
PSE	point of subjective equality
RDK	random dot kinematogram
RM	repeated-measures
RT	response time
s	second(s)
SD	serial dependence

TAE	tilt after-effect
UI	Uniformity Illusion
XCW	counter-clockwise
y/o	years old

Nomenclature

%Higher	In Experiment 6A, proportion of reports wherein test density is deemed higher than the reference density.
%Local	In Experiment 5, proportion of orientation reports in the same direction as the local adapting orientation.
%XCW	In Experiment 5, proportion of 'counter-clockwise' reports.
B	Fixed-effects coefficient estimate.
$\mathbf{BF_{01}}$	Bayes factor for the null hypothesis.
$\mathbf{BF_{10}}$	Bayes factor for the (bidirectional) alternative hypothesis.
$\mathbf{BF_{+0}}$	Bayes factor for the unidirectional (positive) alternative hypothesis.
$\mathbf{BF_{-0}}$	Bayes factor for the unidirectional (negative) alternative hypothesis.
$\mathbf{BF_{Inclusion}}$	Bayes factor for inclusion of a term in a given model.
$\mathbf{BF_{M}}$	Bayes factor of a model.
$\mathbf{C_n}$, $\mathbf{C_{n\text{-}t}}$	In Experiment 3, confidence report in trial n/n-t.

CX In Experiment 5, adapting condition wherein the local Gabor orientation is clockwise and the global pattern orientation is counter-clockwise.

d In Experiment 6, low density (5.77 dots/dva^2).

D In Experiment 6, high density (11.55 dots/dva^2).

dd In Experiment 6, adapting condition with uniform low density across the entire adapting pattern.

Dd In Experiment 6, adapting condition with mixture of central high and peripheral low density.

DD In Experiment 6, adapting condition with uniform high density across the entire adapting pattern.

DIR In Experiments 2B-2E, trials wherein the participant is required to judge the mean direction of the RDK.

dPSE In Experiment 5, the difference between PSE in adapting conditions CX and XC, dPSE = $PSE_{CX} – PSE_{XC}$.

e Euler's number, 2.71828...

E_n, E_{n-t} Absolute value of the response error in trial n/n-t.

J_n Perceptual judgment about the stimulus in trial n, according to the presented model of perceptual decision-making (equivalent to R_n in Experiments 1 – 4).

n	The current trial.
OR	In Experiment 4, Gabor orientation.
R^2	R squared.
R_n, R_{n-t}	In Experiments 1 – 4, response in trial n/n-t.
RAN	In Experiments 2B-2E, trials wherein the participant is required to judge the variance ('randomness') of the RDK.
RE_n, RE_{n-t}	In Experiments 1 – 4, (relative) response error in trial n/n-t.
S_n	Stimulus in trial n.
StD	In Experiments 1 – 3, the standard deviation of the motion directions of the RDK.
t	The number of trial positions away from the current trial n (n-t refers to past trial positions, n+t refers to future trials).
T1	In Experiments 2C-2E, interval between stimulus offset and response cue.
T2	In Experiments 2C-2E, interval between response offset and next trial onset.
V1	Primary visual cortex.

XC	In Experiment 5, adapting condition wherein the local Gabor orientation is counter-clockwise and the global pattern orientation is clockwise.
zRE_n, zRE_{n-t}	In Experiments 1 – 4 normalized response error in trial n/n-t.
σ_R	Response dispersion.

PART I: General Introduction

Abstract

The visual system capitalizes on environmental statistics to efficiently manage and interpret sensory information. This is achieved by compressing visual ensembles into statistical summary representations (ensemble processing) and by informing isolated sensory signals on the basis of the statistical context (contextual modulation of perception). The latter is a critical operation under the predictive coding framework, which regards perception as Bayesian inference. As a measure of the dispersion of visual ensembles over space and time, visual variability is a key descriptor of the environment and crucial to Bayesian computations, informing the precision and reliability of the incoming sensory signal as well as the range of expected options a priori*. In this thesis we examine how the statistical context influences perceptual processing of visual variability. Specifically, we focus in two recently described instances of contextual modulation over time and across the visual field: serial dependence and the Uniformity Illusion, respectively. Both of them are extensively discussed in this General Introduction.*

1. THE NEED FOR EFFICIENT CODING IN THE VISUAL SYSTEM: THE IMPORTANCE OF STATISTICS

The human visual system has a difficult mission: interpreting the external world on the basis of electro-magnetic signals received in the retina and translating these signals into internal representations suitable to be stored, transferred and processed by different neural circuits. This represents managing an enormous amount of information, instantly characterising both familiar and unfamiliar objects in terms of multiple feature-dimensions (spatial location, size, colour, etc) and of their significance for guiding future decisions. Yet this process takes place at almost every instant of our waking life, recreating the complex, ever-changing world in our minds in a seemingly effortless fashion.

All of this is even more perplexing -especially considering how richly detailed our visual experience appears- when we take into account two notable shortcomings of the visual system: it is remarkably limited in informational capacity and generates a large amount of noise at every step of visual processing.

Multiple studies on visual attention (Franconeri, Alvarez, & Enns, 2007; Howe, Cohen, Pinto, & Horowitz, 2010; Pylyshyn & Storm, 1988; Yantis, 1992) and working memory (G. A. Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 2013), as well as phenomena such as change blindness and inattentional blindness (M. A. Cohen, Cavanagh, Chun, & Nakayama, 2012), have shown that only a few items can be attended or remembered at the same time, while most of the detail presented in the visual field goes unnoticed. The striking contrast between the severe capacity limitations demonstrated in controlled experiments and our subjective impression of a rich phenomenology has led some authors to declare that the visual world is nothing but a ‘grand illusion’, barely sustained on perceptual content (Noë, 2002). While others argue that perception is indeed rich, as the described limitations pertain to high-level

processes enabling cognitive access and categorization, which may not capture the entirety of the perceptual content (Haun, Tononi, Koch, & Tsuchiya, 2017a), the fact remains that according to standard paradigms, the amount of visual information available for cognitive processing and decision-making seems severely limited.

As any system of information processing, the visual system is faced with the problem of disambiguating signal from noise. The environment contains external noise due to the stochastic nature of photon emission, but also 'higher-order' noise, i.e. any visual stimulus that interferes with the relevant content in each case –for example, a foggy landscape or ripples on the surface of a pond. But more importantly, the eyes and the nervous system generate a large amount of internal noise. Noise in the visual system arises from multiple sources, including stochastic processes related to cellular activity of photoreceptors and neurons (protein metabolism, vesicle formation), to electrical/membrane activity (channel noise, inducing variations in resting membrane potential and action potential threshold), and to synaptic mechanisms (neurotransmitter diffusion). Furthermore, its effects are amplified by the non-linear and threshold-dependent nature of neural processing: for example, a small variation in the axonal voltage around the threshold for action potential generation can make a fundamental difference in the response of a neural circuit (Faisal, Selen, & Wolpert, 2008).

Given these constraints, it is evident that sensory circuits must be governed by the principle of efficient coding: the neural code (i.e. the transducing function(s) that relate input and response in a single neuron or a neural circuit) must maximize, at every moment, the amount of information about the received input, in face of metabolic demands and internal noise (Attneave, 1954; Barlow, 2012; Chalk, Marre, & Tkacik, 2018). At first consideration, it would seem that the visual system is very poorly equipped to perform this task. How can it overcome such imposing limitations? The answer lies in the redundancy and predictability of the natural world.

The structure of the physical world is full of redundancy and autocorrelation: objects and traits often appear in clusters (such as leaves on a tree) or are aggregated in a predictable fashion - for example, we would not expect to find a wardrobe in the middle of a forest, or an impala grazing in an office room (Kersten, 1987). This autocorrelation occurs both in space and time: the environment is largely stable, with most of the information remaining the same from one instant to the next, and events develop in predictable sequences and frequencies. By exploiting such regularities, the visual system is able to efficiently compress and encode information and make predictions about future events (George A. Alvarez, 2011; Summerfield & Lange, 2014). Indeed, if the environment wasn't predictable perception would be altogether futile, since knowledge of the present state of the environment is only useful to the extent to which it can guide action selection on the basis of predicted outcomes.

In fact, many striking observations about the visual system can be explained in light of statistical optimization. While remarkably inept for detection of specific details, humans and other primates are able to rapidly acquire high-level information about natural images - for example, the presence of a navigable path (Greene & Oliva, 2009) or the emotions prevailing in a large crowd (Haberman & Whitney, 2007). Objects roughly appear the same for a very wide range of light intensities, and we experience the environment as largely stable despite continuous interferences due to blinks, eye movements, etc (Fischer & Whitney, 2014).

Briefly, there are two main approaches whereby environmental regularities can be exploited to optimize efficient coding:

1. By compressing redundant information into summary representations (ensemble processing), so that random uncorrelated noise can be averaged out from the signal (George A. Alvarez, 2011).

2. By informing signal transmission on the basis of the statistical context. For instance:
    a. De-correlating signals maximizes information transmission, according to Shannon's theory of information (E., 1948).
    b. Normalization of the signals enables re-centring the neural code around the most frequent stimulus magnitudes, maximizing its sensitivity within the range of likely inputs (Carandini & Heeger, 2013; Fairhall, Lewen, Blalek, & Steveninck, 2001; Summerfield & Lange, 2014).
    c. Computing the prior probability of occurrence of subsequent stimuli, according to sensory history and context, enables Bayesian computations for interpretation of a noisy input (D. Kersten & A. Yuille, 2003). Some prevailing frameworks regard this as the central operation in perceptual processing: in particular, predictive coding (Clark, 2013) sees perception as Bayesian inference about the state of the world.

In the following sections we will discuss some of the operations that implement the two mentioned approaches to statistics-driven sensory processing (albeit noting that the distinction is somewhat artificial): namely, information compression (ensemble processing) and contextual modulation of perception, which may occur across time (as in sensory adaptation, serial dependence, predictive processing operations, etc) and across the visual space (as in phenomena like perceptual filling-in).

2. ENSEMBLE PROCESSING

In contrast with its poor performance in the detailed representation of individual objects, the visual system is surprisingly good at rapid, accurate computation of summary statistics for visual ensembles, concerning both low and high-level feature-

dimensions, such as size (Ariely, 2001; S. C. Chong & Treisman, 2003), orientation (S. C. Dakin & Watt, 1997), spatial location (George A. Alvarez & Oliva, 2008), brightness (Bauer, 2009), hue (J Maule, Witzel, & Franklin, 2014), motion speed (Watamaniuk & Duchon, 1992), motion direction (Watamaniuk, Sekuler, & Williams, 1989), facial identity (de Fockert & Wolfenstein, 2009), gender and emotional expression (Haberman & Whitney, 2007). These computations often take place in absence of attention and with absent or imprecise representation of the elements of the array (George A. Alvarez & Oliva, 2008, 2009; F. F. Li, VanRullen, Koch, & Perona, 2002).

Extensive research has focused on central tendency statistics, such as the mean, while measures of dispersion, such as variance, have received less attention - see below for a detailed exposition of the research on the latter. In addition to these, other traits, like numerosity (David Burr & Ross, 2008; Steven C. Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011), or higher-order statistics like spatial patterns (M.J. Morgan, Mareschal, Chubb, & Solomon, 2012) or the global properties of a scene (Oliva & Torralba, 2006) are also extracted. Statistical pooling occurs both across space and time: for example, subjects can accurately compute the mean size of a dynamically changing target (Albretcht & Scholl, 2010).

The broad range of dimensions for which summary statistics are computed, the rapidity of the process (as low as 50 ms (S. C. Chong & Treisman, 2003)) and its apparent independence of attention (George A. Alvarez & Oliva, 2008, 2009) and array size (Ariely, 2001; S. C. Chong & Treisman, 2003) have promoted the view of ensemble processing as an automatic, pre-attentive mechanism that precedes the limited capacity bottleneck and plays a fundamental role in visual perception by reducing the computational demands posed by a complex environment. However, some studies paint a slightly more nuanced picture. Investigations based on the simultaneous-sequential method suggest that parallel computation of summary statistics across the same feature-dimension is indeed subject to capacity limitations, similar to parallel perceptual processing of individual items (Attarha & Moore, 2015a). When multiple computations across

different dimensions are performed simultaneously, a study reported that capacity was not compromised (Attarha & Moore, 2015b), while another one found the opposite (Emmanouil & Treisman, 2008). Moreover, sensory prediction error has been found to drive perceptual learning of certain types of summary statistics (such as average location), indicating an influence of top-down tuning signals (Fan, Turk-Browne, & Taylor, 2016). Therefore, it seems that ensemble processing is a highly efficient, but not capacity-free mechanism of information compression, subject to similar principles as other aspects of perceptual processing. Nevertheless, there is no doubt about the computational benefits that it conveys to visual cognition, as reviewed by Álvarez in (George A. Alvarez, 2011):

1. Compressing redundant information of multiple individual items into a single summary representation, critically reducing the computational demands of a resource-limited system.

2. Increasing the precision of the sensory signals (inverse variance of the probabilistic sensory response to a stimulus (Borst & Theunissen, 1999)) and eliminating sensory noise: the average of multiple noisy measures can be much more precise than the individual measures taken separately, as uncorrelated random errors cancel each other.

3. Affording exploration of regions of interest and guiding the focus of attention: by combining individual measures it is possible to obtain spatial patterns of information outside the focus of attention. If visual attention had to be distributed among multiple individual measures, they would be too noisy to be of any use and would not reach awareness - in predictive-coding terms, their low precision would make them unable to update internal models about the world in high-level areas (Feldman & Friston, 2010). Without the ability to form ensemble representations, our visual experience would consist almost only on attended items, and attention would not have any external guide to shift its

focus to regions of interest. Unattended summary representations provide patterns that inform of the type of scene and guide visual search. This is particularly relevant to peripheral vision, which encompasses the great majority of the visual field and where statistics are computed automatically (Balas, Nakano, & Rosenholtz, 2010; L. Parkes, J. Lund, A. Angelucci, J. A. Solomon, & M. Morgan, 2001), informing of regions deserving closer examination by means of foveation (Rosenholtz, Huang, Raj, Balas, & Ilie, 2012).

4. Providing the basis for statistical inference. Ensemble processing enables estimation of population parameters (mean, variance, range) with calculation of confidence intervals, thus enabling classification of objects into groups (set membership), perceptual grouping and detection of outliers.

5. Building a hierarchical representation of the scene, with integration across multiple levels of abstraction. This is important, for instance, for perception of textures (S. C. Dakin & Watt, 1997; Michael J. Morgan, Chubb, & Solomon, 2014; M.J. Morgan et al., 2012) and identification of scene types (F. F. Li et al., 2002; Oliva & Torralba, 2006). It also enhances the capacity of visual working memory compared to retention of unrelated items (Timothy F. Brady & Alvarez, 2015).

What are the mechanisms responsible for ensemble processing? A prevailing notion, especially in early research, points to the distribution of attention among all the objects of the array and computation of summary statistics from all individual representations, which are subsequently discarded (as proposed by Ariely in (Ariely, 2001) and further elaborated in (George A. Alvarez, 2011)). Although several studies appear to support this suggestion (Ariely, 2001, 2008; S. C. Chong & Treisman, 2003, 2005a, 2005b), it has also provoked scepticism, since such mechanism would require parallel processing of a number of items exceeding the capacity limits of visual attention, especially for elements involving high-level areas such as faces (Pylyshyn & Storm, 1988). Consequently, some models propose a strategic subsampling of a few items that could

account for the accuracy reached in several experiments (Myczek & Simons, 2008). For example, a study on variance computation reported a very inefficient mechanism, with only 5-6 items being used, of an array formed by more than 100 (M.J. Morgan et al., 2012). However, this notion of a 'strategic subsampling' has also been contested by different studies, such as the ones performed with crowded ensembles -where individual items are not distinguishable- (L. Parkes, J. Lund, A. Angelucci, J. Solomon, & M. Morgan, 2001) or with different conditions that would demand different strategies, if strategic selection was the norm (S. Chong, Joo, Emmanouil, & Treisman, 2008). Finally, other authors consider that individual representations are not needed at all (Jennifer E. Corbett & Oriet, 2011), and ensembles are directly processed in a manner akin to textures: that is, performing overall feature processing (filled surface, spatial frequency, density), in absence of object segmentation (Balas et al., 2010; Im & Halberda, 2013; Tibber, Greenwood, & Dakin, 2012).

Even the question of whether there is a single, domain-general system supporting ensemble processing or widespread domain-specific operations remains unanswered. As described above, there is conflicting evidence about whether (Emmanouil & Treisman, 2008) or not (Attarha & Moore, 2015b) simultaneous computation of summary statistics across different feature-dimensions is subject to interference -an issue that would help to resolve the question. In this regard, by using an individual differences approach, a study reported lack of correlation between subject´s performance in ensemble processing of high-level versus low-level feature domains, suggesting that they are supported by independent functional architectures (Haberman, Brady, & Alvarez, 2015). According to this, ensemble processing may be a ubiquitous, canonical neural computation performed independently on different feature-dimensions. On the other hand, another study (Zhao, Ngo, McKendrick, & Turk-Browne, 2011) found bidirectional interference between ensemble perception and statistical learning, suggesting a degree of commonality in the underlying mechanisms involved in computation of summary statistics for an ensemble and acquisition of statistical

regularities (co-occurrence of individual elements) across sequentially presented ensembles.

Though little is known about the brain structures responsible for generating summary representations, a role has been proposed for the ventral stream, a sequence of cortical areas beginning at the primary visual cortex (V1) and extending to the inferotemporal cortex (IT), which compresses sensory information into patterns with progressive loss of local detail in favour of increasing complexity in downstream areas (Freeman & Simoncelli, 2011).

Aiming to identify the specific area responsible for ensemble processing, a study (Freeman & Simoncelli, 2011) employed perceptual metamers: modified pictures that preserve the appearance of the original image by removing peripheral detail while maintaining the summary statistics within pooling regions, thus mimicking the loss of visual detail that happens along the ventral stream. Neural receptive fields (areas responsible for pooling of the information of a certain region in the visual field) increase in size across retinal eccentricity (Freeman & Simoncelli, 2011; Rosenholtz et al., 2012) and along the ventral stream. The rate of increase across eccentricity is larger for downstream areas, which provides a signature for identifying the area responsible for a specific process. By analysing the rate of compression suitable for maintaining identical appearance, this study identified V2 as the likeliest candidate for the generation of summary representations.

Another fMRI investigation presented evidence of ensemble-related neural adaptation driven by different stimulus types in the anterior-medial ventral visual cortex, an area involved in texture and scene processing (Cant & Xu, 2012), also part of the ventral stream.

In visual perception, the ventral stream operates independently of an alternative pathway devoted to detailed object processing, supported by lateral occipital and

parietal areas (Cant & Xu, 2012). The degree to which each region of the visual field partakes in these two processing systems in vastly inhomogeneous, with the central 2° (fovea) equipped with a powerful anatomical machinery for detailed resolution (photoreceptor density and type, ganglion cell density, receptive field size, etc) and the vast majority of the visual field (periphery) pooling low spatial frequency signals into summary statistics. Indeed it has been claimed that peripheral vision is entirely statistical, and mathematical models based on this assumption have been able to predict crowding effects (Balas et al., 2010), visual search and discrimination tasks (Rosenholtz et al., 2012) and to capture texture appearance (Freeman & Simoncelli, 2011) (but see (Alexander, Schmidt, & Zelinsky, 2014)). Interestingly, there seems to be a lot in common between ensemble processing and the characteristic peripheral phenomenon of crowding (wherein discrimination of the traits of a peripheral stimulus is impaired if there is another object nearby): both of them are based on pooling of visual information with loss of local particularities (Balas et al., 2010), both have been related to texture processing (L. Parkes et al., 2001) and seem to occur independently in different brain areas for different feature-dimensions (Whitney & Levi, 2011).

Considering all the above, the resolution of the conflict between the poor capacity of the visual system and our subjective rich perceptual experience may not be a 'grand illusion', but rather the result of the integration of coarse peripheral summary statistics with the detail provided by foveal vision (Michael A. Cohen, Dennett, & Kanwisher, 2016) (although some degree of inflation, not sustained of perceptual content, is likely to contribute to peripheral appearance (Odegaard, Chang, Lau, & Cheung, 2018; Solovey, Graney, & Lau, 2015)). How this foveal-peripheral integration occurs is as yet not clearly established, but likely relies on contextual modulation of imprecise sensory signals by visual information received in recent history or other regions of the visual field. In later sections we will review some of the mechanisms supporting this contextual modulation, which are the focus of our research.

2.1. VISUAL VARIANCE PROCESSING

2.1.1. Early works on variance in Psychophysics

Compared with central tendency measures (e.g. mean), less attention has been paid to how the brain processes dispersion statistics, particularly variance, even if visual variability is a fundamental descriptor of the environment (Michael, Gardelle, & Summerfield, 2014).

Still, there are some interesting reports in early psychophysical research, as reviewed by Peterson in (Peterson & Beach, 1967). Several researchers (Lee Roy Beach & Swenson, 1966; Spencer, 1961) examined how variance modulated judgments about means and found that the variance of estimates increased with the variance of the sample, though it didn't bias the average results. It was proposed that subjects were actually making inferences about the mean as a parameter, with sample variance influencing the standard error of the mean. As for explicit judgments of variability, the matter used to receive more attention decades ago than in the recent past. In 1939, Hofstatter (Hofstatter, 1939) observed a positive correlation between stimulus variance and subjects' variance estimates, but a negative one between stimulus mean and subjective variance, as if the participants were judging relative discrepancies in proportion to the mean, an interpretation related to the coefficient of variation and Weber's law in psychophysics. Subsequent studies (Lathrop, 1967) replicated this finding. Also, rather than the mathematical variance (with square distances from the mean) or even the coefficient of variation, the power assigned to the distances from the mean to estimate global variability seemed to vary in relation to the context. When experimental conditions emphasized large deviations, large exponents fitted better the subjects' responses; when small deviations were emphasized, it was the opposite (Hofstatter, 1939). Using normal distributions, Beach and Scopp (L.R. Beach & Scopp, 1967) found that a small power was a better fit for their data, which could be expected given the

clustering of normally distributed data around the mean. They hypothesized that this could be different for other types of probability distributions.

2.1.2. Indirect approaches: modulation of perceptual processes by stimulus variance

Modulation of accuracy and confidence in mean judgments

The modulation of mean estimates by variance has returned to the spotlight in recent times. Several studies have reported longer reaction times and worse accuracy in averaging judgments in presence of a greater variability in decision-relevant evidence, concerning hue, shape (Gardelle & Summerfield, 2011), motion direction (Gardelle & Mammasian, 2015) and orientation (Zylberberg, Roelfsema, & Sigman, 2014). Likewise, it has been found that a higher variance, even for equal range, impaired hue averaging (John Maule & Franklin, 2015) and that subjects were less likely to find the mean hue familiar in a membership identification task when hue variance increased (J Maule et al., 2014). By asking subjects to determine which of two groups of bars had a greater average length, another team found, again, that higher variance impaired performance, and that subjects behaved as if they were judging mean difference by the Student's T test, assigning confidence limits close to those employed in formal statistics, with a one-tailed significance of P=0.05 (Fouriezos, Rubenfeld, & Capstick, 2008). Evidence also suggests that there is a limit to the amount of variance that can be compressed in a summary representation (John Maule & Franklin, 2015) and that variance influences perceptual grouping and working memory capacity (Timothy F. Brady & Alvarez, 2015).

Thus, mean judgements are subject to modulation by variance (external noise), as described by models of perception that represent encoding of ensembles and individual items in the form of a probabilistic distribution or a set of probability distributions, each subject to internal noise (George A. Alvarez, 2011; Ma, 2012). Greater variance between the individual items in the ensemble translates into greater variance in the internal

representation of the mean (and therefore larger errors and lesser consistency in subjects' responses (Zylberberg et al., 2014). Models of mean size (Gardelle & Summerfield, 2011) and orientation (V. Li, Herce Castanon, Solomon, Vandormael, & Summerfield, 2017; Zylberberg et al., 2014) judgments achieve a better fit for describing subjects' performance when taking both stimuli variance and other sources of noise into account.

In light of all this, several authors regard variance as a measure of the reliability of sensory evidence for perceptual decisions (Gardelle & Mammasian, 2015; Gardelle & Summerfield, 2011; Spence, Dux, & Arnold, 2016; Zylberberg et al., 2014), which is related, in psychophysical terms, to the signal-to-noise ratio and the slope of the psychometric curve (Gardelle & Mammasian, 2015), and in Bayesian terms to the width of the likelihood distribution. Whether and how it has an impact in subjects' confidence in such decisions is still under discussion.

In a study on average orientation judgments (Zylberberg et al., 2014), higher stimulus variability was systematically accompanied by lower accuracy but higher confidence. The authors explained this counter-intuitive result by subjects' inability to properly adapt their decision and confidence criteria to the properties (variance) of the stimulus. Like similar observations regarding paradoxical effects of internal noise (D. A. Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012) and attention (D. Rahnev et al., 2011), this finding may be explained by a model (D. A. Rahnev et al., 2012) based on signal-detection theory (Green & Sweets, 1966), with the critical constraint of a unique set of decision and confidence criteria used by the observer across all experimental conditions (Gorea & Sagi, 2000) -or, in a less strict formulation, an insufficiently adaptable set of criteria (Zylberberg et al., 2014). Due to the probabilistic nature of sensory processing, a noisier signal (for example, a high-variance ensemble) is more likely to generate a highly deviant internal response (very low or very high). If the observer fails to adapt their criteria to the amount of noise on a trial-by-trial basis, this would mean that noisier trials would produce more ratings (inaccurate yet) well above the high-confidence

criterion threshold. In this light, the inability to tune decision and confidence to the ensemble variance would stem from subjects' underestimation of the variance of the sensory evidence or a suboptimal application of this variance to judgments of stimulus reliability (Zylberberg et al., 2014), an issue that has been previously reported in behavioral economics (Tversky & Kahneman, 1974).

On the other hand, by requesting judgments about average motion direction, another study reported that variance undermines confidence even to a greater extent than accuracy/sensitivity (Spence et al., 2016). The authors discussed that confidence judgments were disproportionally affected by the variability of neural responses in perceptual areas, i.e. the width of the response function of the neural population, in accordance with Bayesian accounts of confidence as a measure of the precision of the perceptual-decision processes (Meyniel, Sigman, & Mainen, 2015).

A third study on the relationship between ensemble variance and confidence in mean judgments might shed light on the discrepancy between the previous two. This research work (Gardelle & Mammasian, 2015) independently manipulated mean and variance of random dot kinematograms (RDKs) for a motion categorization task. After matching performance between conditions, confidence reports in low-mean (mean close to category boundary) and low-variance (high precision) RDKs were compared with high-mean high-variance stimuli. Confidence reports, which positively correlated with performance, were influenced by stimulus variability in an idiosyncratic manner: some subjects placed more weight, in terms of confidence, in the distance to category boundary (mean), while others were more sensitive to stimulus variance; this preference remained stable for each individual in a second session one week later. Besides, those subjects whose confidence was more affected by variance were also those whose performance was particularly impaired by variance, and vice versa. The authors proposed that some participants may not need to maintain a unique set of criteria throughout the entire experiment; on the contrary, some may have a greater ability to adjust their decision and confidence criteria to each trial's level of internal

noise, and therefore their confidence judgments may be affected by the individual´s ability of estimating current and prior noise variance or, in a more general perspective, by a misperception (under or overestimation) of statistical variability.

In summary, the disparate evidence regarding variance and confidence appears to depend on the accuracy or weight of variance perception, an aspect of ensemble processing that demands closer examination -see below for an account of the studies on direct variance estimates (section 2.1.3. of the current chapter).

Modulation of neural adaptation and priming

The influence of variance has been studied for other processes dependent on statistical context, such as adaptation and priming. The variance of sensory input across time modulates neural adaptation: in low-variance environments, sensory signals attain higher resolution/sensitivity, both in the visual and auditory domains (Dahmen, Keating, Nodal, Schulz, & King, 2010; Fairhall et al., 2001). Exploring adaptation aftereffects in an ensemble-processing task, Whitney and colleagues found that adaptation to mean size was stronger for low-variance ensembles, i.e., when the precision of the mean estimate was higher (Jennifer E. Corbett, Wurnitsch, Schwartz, & Whitney, 2012). E. Michael and colleagues described an effect of priming by the variability of visual information (Michael et al., 2014), by which judgements about the mean of an array were faster and more accurate if preceded by another array with the same variance, independently of mean values and overlapping of individual elements. Results suggested that variance priming was independent of attention, but occurred through feature-dimension specific channels. The effect was visible with prime-target intervals as short as 100 ms, pointing to an automatic, pre-attentive mechanism. To summarize these results, the authors explained the effect of variance priming in terms of adjusting the gain of the sensory signals to fit the statistics of the environment.

In summary, contextual variance provides a measure of the expected range of stimuli in the environment (in Bayesian terms, the width of the prior distribution) (Michael et al., 2014), and therefore it is a key factor to the tuning of perceptual prediction.

2.1.3. Direct approaches: explicit variance estimation

Compared to the effect of variance on ensemble perception, the direct examination of variance as a distinct perceptual feature has received less attention in recent research.

One study (Kareev, Arnon, & Horwitz-Zeliger, 2002) set out to address the long-standing claim (made even by Francis Bacon in 1620 (Bacon, 1854)) stating that humans perceive the world as more regular, i.e. less variant than it actually is. The authors reasoned that such misperception, if real, could be due to the use of sample variance as an estimate of population variance without correction by sample size, and verified this hypothesis in a series of experiments requesting estimates of visual ensemble variance.

In the context of research on texture discrimination, another team examined perception of pattern regularity as a summary statistic, given by the positional variances of each element of the pattern with respect to the ideal position for an absolute regular pattern (equal distance between homologous elements of the pattern) (Michael J. Morgan et al., 2014; M.J. Morgan et al., 2012). The authors reasoned that statistical computation would explain why regular patterns appear regular despite the expected distortions on the perceived position of individual elements produced by sensory noise. Pattern discrimination was more sensitive when neither of the two patterns was perfectly regular, explained by a pedestal effect caused by internal noise: a low non-zero variance would raise the response of the mechanism just below the threshold for discarding noise, so that an additional increment would be easily detected. As pedestal variance increased, the 'just noticeable difference' (JND) in positional variance first decreased and then raised again, forming a 'dipper function', resembling those observed in

contrast or blur discrimination (Michael J. Morgan et al., 2014): thus, except for very small non-zero amounts of variance, as this statistic increased, detecting further irregularities became increasingly harder. In this sense, the authors regarded camouflage effects as masking of (relevant) variance by (irrelevant) variance. Besides, results suggested a very inefficient mechanism of variance computation, with subsampling of only 5-6 elements out of over 120.

While very different in their aims and methods, the two mentioned studies (on variance underestimation (Kareev et al., 2002) and pattern regularity) both suggest that the visual system conflates external visual variability and internal noise to some extent, smoothing away variance of the sensory signals unless there is strong evidence about it being a physical trait of the environment -for example, if it raises above a certain threshold.

Another study employed judgements of colour diversity (or variability) in colour ensembles to investigate whether perceptual phenomenology was indeed richer that cognitive access to perceptual contents -in other words, to assess the relationship between phenomenological and access consciousness (Zohar Z. Bronfman, 2014). Colour diversity could be estimated without attention, and did not impair performance on a different, attended task. Irrelevant diversity affected judgments on relevant diversity. Masking impeded diversity estimations, implying that such judgement cannot be made without awareness. The authors concluded that colour diversity estimation was an automatic, pre-attentive mechanism which did not consume significant cognitive resources, yet it appeared to require a conscious representation (albeit transient) of the individual colours; thus, results cautiously supported that phenomenological consciousness is based on rich perceptual content even if only a small fraction can be accessed by visual attention and cognitive resources.

In a similar vein, a research work asserted that variance judgments required processing of the individual elements of the ensemble, in this case concerning a high-level trait such as facial emotion (Haberman, Lee, & Whitney, 2015). Variance estimation was greatly

impaired in conditions that typically hindered high-level facial processing, like in upside-down face ensembles.

Another team (Payzan-LeNestour, Balleine, Berrada, & Pearson, 2016) reported negative adaptation after-effects on variance judgments of diverse types (motion variance, abstract visual calculations) which generalized across the different representations. This result suggests that variance adaptation is not of sensory origin but operates at a high (cognitive) level of processing, and, therefore, that variance constitutes an abstract property independent on the dimension about which it was initially computed.

3. CONTEXTUAL MODULATION OF PERCEPTION

Considerable evidence indicates that perception (and vision in particular) is affected by previous stimulus history, as well as by stimuli presented in other regions of the visual field or other sensory modalities. Henceforth we will refer to this influence as contextual modulation of perception, and will focus on some recently described instances regarding temporal and spatial visual context.

As previously described, the visual system has evolved to cope with the vast complexity of the natural world and its own limitations by exploiting temporal and spatial autocorrelations in the environment. Given that the visual context is informative for the interpretation of a single, noisy sensory signal, one may surmise that the combination of both inputs may be a basic operation in perceptual processing. This intuition can be formalized in Bayesian terms, where the sensory input takes the form of a likelihood distribution (the stimulus is encoded as a probabilistic response by a neural population) and the prior distribution encodes the probability of encountering each possible

stimulus given the available (contextual) information. In other words, the brain exploits contextual information to build up expectations for incoming stimuli. Numerous studies have indeed encountered pre-stimulus changes in neural activity driven by such expectations, in sensory, decision and motor areas involved in perceptual decision-making, as reviewed extensively by Summerfield and colleagues (Summerfield & Egner, 2009; Summerfield & Lange, 2014).

Predictive coding is a comprehensive framework that aims to capture this purported Bayesian basis of perception, along with some notable properties of the anatomical organization of the brain and some well-studied physiological and psychophysical phenomena. This framework proposes that perception arises as a result of hierarchical Bayesian inference, whereby information of increasing complexity and abstraction is processed by a multi-level hierarchy of brain areas. At each level, the noisy signal received through the senses or from lower layers of the hierarchy (likelihood) is combined with a prediction (prior) generated at higher levels that operate at a broader scope and thus provide informative context that feeds back to lower areas (Clark, 2013; K. Friston, 2005; Kersten, Mamassian, & Yuille, 2004; Rao & Ballard, 1999). Predictive coding offers a plausible interpretation for the increasing size and scope of the receptive fields in downstream cortical areas (a prominent example being the ventral stream) (K. Tanaka, 1996), and the structure of anatomical connections across different areas, where backward connections frequently outnumber forward connections (Angelucci et al., 2002). This framework is also able to explain phenomena such as repetition suppression (Grill-Spector, Henson, & Martin, 2006), and extra-classical effects modulated by information outside the neurons' receptive field, which cannot always be accounted for by transmission through lateral connections (Rao & Ballard, 1999).

Despite the merits of the Bayesian framework, it is evident than a model of perception based on strict Bayesian computations would be too simplistic and inaccurate. On the one hand, it would obviate the fact that some of the most widespread phenomena related to contextual modulation of perception, namely adaptation after-effects, are

seemingly anti-Bayesian, i.e. repulsive (or negative) biases *away* from previous history. On the other hand, a purely Bayesian model would need to account for all the possibilities *a priori*, which would be computationally intractable. As described above, perceptual processing must be constrained by the principle of efficient coding, to optimize resource allocation to the likeliest possibilities. Sensory adaptation plays a crucial role in achieving this goal, by normalization of neural activity with respect to a broader context (Carandini & Heeger, 2013), thus re-centring the neural code and maximizing sensitivity around the likeliest stimuli (Fairhall et al., 2001; May & Zhaoping, 2016; Webster, 2011). In this regard, a Bayesian observer model of perception constrained by the principle of efficient coding (with the prior and likelihood distributions modulated by prevailing stimulus statistics) has been shown to reproduce seemingly anti-Bayesian (repulsive) effects (Wei & Stocker, 2015). Thus, environmental statistics are fundamental for an efficient application of Bayesian operations in perceptual processing: in particular, visual variability (or variance) sets the range of likely stimulus magnitudes, namely the width of the prior distribution, and tunes neural responses accordingly (Dahmen et al., 2010; Fairhall et al., 2001; Michael et al., 2014).

Furthermore, not all signals are equally reliable. In Bayesian processing, input signals are weighted by their precision to minimize the effect of external and internal noise, and thus adjust the degree of reliance in prior history according to the quality of current input. Within the predictive coding framework, attention is deemed a mechanism for tuning the precision (inverse variance) of the sensory/prediction error signals (Feldman & Friston, 2010), and metacognitive confidence is thought to arise from an estimate of that precision (width of the distribution of neural activity elicited by a signal – the likelihood function) (Drugowitsch & Pouget, 2012; Meyniel et al., 2015). As an estimate of dispersion, environmental variance provides a measure of the precision of the summary representations that form a large part of our perceptual experience (Jennifer E. Corbett et al., 2012). Thus, variance appears, in different ways, as a crucial statistic in the Bayesian-based interpretation of the natural world.

Our research examines two recently described examples of contextual modulation of perception that can be interpreted on the basis of Bayesian-like operations, with a prior constructed through time (serial dependence) or across the visual field (the Uniformity Illusion), and the way visual variability is processed through these two phenomena. We will now briefly describe the basic properties of these processes.

3.1. CONTEXTUAL MODULATION OF PERCEPTION THROUGH TIME: SERIAL DEPENDENCE IN PERCEPTUAL DECISION-MAKING

It has been known since Antiquity that perception is subject to history-dependent biases. Some of the most striking and widely studied are adaptation after-effects, whereby (relatively long) exposure to a certain stimulus magnitude will bias subsequent perception away from that magnitude (negative, repulsive effect) (Addams, 1834; Aristotle, 350 BCE; Gibson & Radner, 1937).

Comparatively, positive (attractive) biases have received less attention. Yet it has also been known for long that perceptual decisions are attracted toward the mean of the distribution of past stimuli as well as toward the most recent presentations. These two effects, the first one termed regression effect or contraction bias (first described by Hollingworth in 1910 (Hollingworth, 1910; Stevens & Greenbaum, 1966; Teghtsoonian & Teghtsoonian, 1978)), and the second one termed sequential effect (Collier & Verplanck, 1958; DeCarlo, 2006; Mori, 1998; Senders & Sowards, 1952; Yu & Cohen, 2008), serial dependence (Fischer & Whitney, 2014; Fründ, Wichmann, & Macke, 2014) or recency bias (Kalm & Norris, 2017), were traditionally studied as separate phenomena. Crucially, these attractive effects to previous history arise even when past stimuli are non-informative about the current state, such as in random sequences of stimuli, as seen in typical psychophysical experiments (Fründ et al., 2014; Yu & Cohen, 2008). Indeed, part of the reason why the recency bias in perception has been largely

overlooked might be that it is usually more convenient to assume that responses to individual trials in a perceptual task are independent from each other (Fründ et al., 2014).

This situation of relative neglect has started to change in recent years. In 2014, Fischer and Whitney published an extensive quantitative analysis on serial dependence in visual perception (Fischer & Whitney, 2014). In a series of experiments requiring successive judgments on the orientation of supra-threshold Gabor patches, these authors described a strong and systematic bias of visual judgments toward recent stimuli, which was tuned by spatiotemporal proximity and stimulus similarity, gated by attention and observed even in absence of a motor response or explicit memory of the past stimulus. Subsequently, the same team reported serial dependence for spatial position (Manassi, Liberman, Kosovicheva, Zhang, & Whitney, 2018), complex objects (face identity, even across different viewpoints (Alina Liberman, Fischer, & Whitney, 2014)), abstract judgments (attractiveness (Xia, Liberman, Leib, & Whitney, 2015), emotional expression (A. Liberman, Manassi, & Whitney, 2018)) and statistical properties (ensemble mean (Manassi, Liberman, Chaney, & Whitney, 2017), numerosity (J. E. Corbett, Fischer, & Whitney, 2011)), showing that serial dependence is not restricted to low-level visual features. Other authors have expanded the knowledge about the phenomenon on a wide variety of visual dimensions, and provided insight on its properties, mechanistic basis and biological function (Alais, Lelung, & Burg, 2017; Bliss, Sun, & D'Esposito, 2017; Cicchini, Anobile, & Burr, 2014; Cicchini, Mikellidou, & Burr, 2017; Fritsche, Mostert, & Lange, 2017; John-Saaltink, Kok, Lau, & Lange, 2016; Kok, Taubert, Burg, Rhodes, & Alais, 2016; Taubert, Alais, & Burr, 2016). Likewise, Lau and colleagues described a similar attractive bias involving metacognitive (instead of perceptual) representations, reporting an intertask ‘confidence leak’ that could not be explained by priming or fluctuations in attention (D. Rahnev, Koizumi, McCurdy, D'Esposito, & Lau, 2015).

The wide range of dimensions that exhibit serial dependence suggest that this phenomenon has evolved as a critical trait to the optimization of perceptual

mechanisms. Strikingly, however, serial dependence may lead to *suboptimal* performance in perceptual decisions -that is, in standard laboratory tasks, when sequential presentations are statistically independent. Fischer and Whitney (Fischer & Whitney, 2014) reported a *decreased* sensitivity to slight changes in orientation, driven by the influence of a previous similar tilt, which could be measured by a less pronounced slope of the Psychometric function relating orientation and judgment. This sets serial dependence apart from a superficially similar phenomenon like priming, whereby perceptual judgments about a stimulus exhibit *increased* accuracy and sensitivity if preceded by the same stimulus or by another sharing similar properties (Maljkovic & Nakayama, 1994; Posner & Snyder, 1975). The apparent disadvantage conveyed by serial dependence prompts the question of its biological function. As with many visual illusions, this phenomenon may seem maladaptive in laboratory conditions, but it is actually revealing of how the evolution of the visual system has been guided by the peculiarities of the natural world. The environment is largely stable, and vision has evolved to perceive it as such, in spite of disruptive events such as blinks, eye movements or momentary occlusions of objects, and random fluctuations in sensory signals due to internal noise. As Fischer and Whitney proposed (Fischer & Whitney, 2014), serial dependence would promote perceptual stability and smooth away noise in visual input, at the cost of reducing sensitivity to small, transient variations. These authors proposed the term 'continuity field' to describe the spatiotemporal operator over which perception is attracted to past input; they also proposed that, for high-level objects (such as faces), the continuity field might be object-centred, instead of defined by spatiotemporal coordinates, in order to facilitate object-invariance (Alina Liberman et al., 2014). Similarly, Lau and colleagues proposed a metacognitive continuity field that would explain 'confidence leak', justified by the stability of environmental noise affecting the quality of the perceptual signal, and supported by a common confidence 'currency', independent of stimulus and task (D. Rahnev et al., 2015).

In light of this proposed biological function, serial dependence and sensory adaptation would be complementary mechanisms, generating opposite biases in different

timescales in order to handle the necessary balance between perceptual stability and sensitivity to change: serial dependence would act on transient fluctuations, assumed to be based on internal changes, while sensory adaptation would prevail after more prolonged exposure, likely to represent an external change in environmental conditions (Fischer & Whitney, 2014; Manassi et al., 2017). In this vein, a study examined serial dependencies on two different traits of a single stimulus (faces), specifically one stable trait (gender) and another changeable (emotional expression), and reported positive serial dependence for the stable trait and negative effects for the changeable one, thus highlighting different coding strategies for various properties of the same object, in function of expectations about their stability (Taubert et al., 2016).

The incorporation of prior knowledge to inform decisions in face of uncertainty (noise) strongly echoes the dynamics of Bayesian decision theory. This notion is consistent with the prevailing view of perception as Bayesian inference within the predictive coding framework (Daniel Kersten & Alan Yuille, 2003; Maloney & Mamassian, 2009; Petzschner, Glasauer, & Stephan, 2015; Yuille & Bülthoff, 1996), and indeed, Bayesian observer models have been used by several authors to represent perceptual biases (Jazayeri & Shadlen, 2010; Kording & Wolpert, 2004; Petzschner & Glasauer, 2011). Concerning prior knowledge, it is reasonable for a Bayesian observer to assign more weight to more recent inputs, which are usually better predictors of the current state of the environment (Anderson & Schooler, 1991; Kalm & Norris, 2017). This would require the Markovian implementation of a dynamically changing prior that continuously incorporates information about the most recent input, in a process akin to a Kalman filter (Petzschner & Glasauer, 2011). Such prior would contain information about the whole stimulus history and, as more stimuli are supplied to the observer, would tend to approximate to the statistical distribution of the stimuli in the environment, weighted by recency of presentation. In this regard, models based on an iteratively updating prior have proven superior to alternative models incorporating a fixed prior based on the stimulus statistics (Kalm & Norris, 2017; Petzschner & Glasauer, 2011). Importantly, such models are capable of reproduce not only recency biases (serial dependence), but also

regression effects (Petzschner & Glasauer, 2011; Raviv, Ahissar, & Loewenstein, 2012), given the attraction exerted by the prior toward the centre of the distribution of stimuli. A common process is thus able to explain two different history-dependent effects operating on different timescales.

What are the specific processes and brain areas that give rise to serial dependence? In their influential 2014 study, Fischer and Whitney asserted the perceptual origin of the bias (as opposed to a decisional, or memory-based attraction), as serial dependence was observed in a two-alternative forced-choice (2AFC) task, a result hard to explain by a shift in a decision criterion (Fischer & Whitney, 2014) -note, however, that the sample size was only 3 subjects. These authors also proposed that a model based on activity-dependent gain changes or tuning shifts in sensory neurons could explain their pattern of results. In a similar vein, St. John-Saaltink and colleagues reported that orientation-selective fMRI activity signal in the primary visual cortex (V1) was biased to the orientation of the previous trial (John-Saaltink et al., 2016), suggesting a neural correlate of serial dependence in sensory areas. Importantly, these authors reported that the encountered bias in V1 activity correlated with the previous perceptual choice, and not the bottom-up sensory signal -in other words, it was related to the previous response, and not the previous stimulus. However, technical limitations in the temporal resolution of fMRI must be considered when assessing these results (as the reported bias could be residual BOLD activity from the previous trial), even though the authors made provisions about this issue. In other two behavioural studies on orientation (Cicchini et al., 2017) and position-related (Manassi et al., 2018) serial dependence, reported response patterns were, according to authors, consistent with a perceptual origin: specifically, Cicchini and colleagues found positive serial dependence in a 2AFC task, particularly when the perceptual distance between stimuli was small and sensory reliability was low (Cicchini et al., 2017), whereas Manassi et al. observed serial dependence effects at the time of perception (Manassi et al., 2018). Finally, another team (M. Fornaciai & Park, 2018) reported neurophysiological signatures of serial dependence (along with

behavioural results) in absence of an explicit task, again supporting that serial dependence arises from perceptual processes.

Other authors, however, have presented evidence supporting the opposite hypothesis, namely serial dependence being a bias on perceptual decisions exerted on (and by) working memory representations, in absence of alterations in perceptual content. Bliss and colleagues found that serial dependence in spatial position was absent at the time of perception but increased if a delay was enforced between stimulus offset and response, indicating that the bias arose from the maintenance of visual representations in working memory, where they were distorted by previous representations (Bliss et al., 2017). Fritsche and colleagues (Fritsche et al., 2017) found that positive serial dependence in orientation only occurred in tasks liable to decisional biases (adjustment task), while a negative tilt after-effect was observed in perceptual comparison (2AFC and equality) tasks, more suitable to the isolation of purely perceptual effects. Besides, they also observed an increase in serial dependence with delayed responses (in the adjustment task), in the same vein as Bliss's findings. They concluded that sensory representations were subject to negative adaptation effects while decisions were biased toward recent information in working memory, as a joint system to balance sensitivity and stability. Alais et al. (Alais, Lelung, et al., 2017) reported convergent results, observing linear summation of concomitant negative and positive dependencies influencing motion perception (wherein motion judgments were subject to positive decision biases and orientation reports experienced negative effects). Nevertheless, these conclusions have been contested by similar studies from other teams (Cicchini et al., 2017; Manassi et al., 2018), as reported in the previous paragraph.

Although both hypotheses about the basis of serial dependence (i.e., perceptual and post-perceptual) have several pieces of evidence in their favour, the decision-memory account is better linked to extant knowledge on adjacent fields, especially on the properties of visual short-term memory. A large bulk of scientific evidence shows that short-term memory representations are biased by previous history in the visual

(Ashourian & Loewenstein, 2011; Huang & Sekuler, 2014; Olkkonen, McCarthy, & Allred, 2014; Papadimitriou, Ferdoash, & Snyder, 2015), auditory (Akrami, Kopec, Diamond, & Brody, 2018; Lockhead & King, 1983; Lu, Williamson, & Kaufman, 1992; Visscher, Kahana, & Sekuler, 2009) and tactile (Fassihi, Akrami, Esmaeili, & Diamond, 2014; Preuschhof, Schubert, Villringer, & Heekeren, 2010; Romo & Salinas, 2003) domains. This dovetails with an ongoing debate between two different models about the nature of working memory: one that sees it as a system capable of storing a discrete amount of independent representations in a limited number of fixed slots (slot-based model) and another that considers working memory as a 'fluid' resource that can be flexibly allocated to maintain different types of representations, including simple features, high-level objects or hierarchical ensemble representations (resource-based model) (Suchow, Fougnie, Brady, & Alvarez, 2014). The resource-based model allows for a much more natural interpretation of the interaction of sequential (or even simultaneous (Huang & Sekuler, 2014)) memory representations, with newer representations partially overwriting previous ones in a kind of leaky-integration process (Kalm & Norris, 2017; Matthey, Bays, & Dayan, 2015), as well as the enhancement of the recency bias in presence of capacity-related constraints for the storage of new information (Huang & Sekuler, 2014). Such a process allows for accumulation of noisy evidence into complex, even multi-hierarchical representations (Timothy F. Brady & Alvarez, 2015), facilitating the sequential acquisition of information about complex or imprecise objects (Busemeyer & Myung, 1987). Ultimately, this continuous updating of flexible memory representations leads to the generation of an iterative prior that contains information about all past states, and represents a summary of the statistics of the presented stimuli, weighted by recency; thus, it is consistent with Bayesian dynamics and able to explain both regression effects and serial dependence (Kalm & Norris, 2017; Olkkonen et al., 2014; Raviv et al., 2012). A recent study has provided a plausible neural basis for this process, reporting that the posterior parietal cortex (PPC) stores information about stimulus history and mediates attractive biases on perceptual decisions (Akrami et al., 2018).

Still, the debate remains about whether serial dependence is entirely explained by a bias in working memory. Noting the similarities between serial dependence in perceptual tasks and some well-studied phenomena in working memory tasks (particularly proactive interference), Kiyonaga, Whitney and colleagues have made an attempt to bring together the available knowledge from both research fields (A. Kiyonaga, J. M. Scimeca, D. P. Bliss, & D. Whitney, 2017). They proposed that history-driven attraction may be a widespread process to promote stability and cohesion in different types of internal representations, not only perceptual, but also cognitive. Thus, serial dependence may arise at different levels of processing, including perception, attention and memory. In this regard, memory representations of low-level features have been described in early sensory areas, possibly even V1 (Baumann, Endestad, Magnussen, & Greenlee, 2008), prompting the suggestion that a memory-driven bias might arise at a very early stage of perceptual processing. In a similar vein, in higher-order visual areas, iconic representations are generated to promote object invariance (Podvalny et al., 2017). Nevertheless, for the moment the degree of commonality of serial dependence and visual working memory, and in broader terms, the neural basis of serial dependence, remains an open question.

3.2. CONTEXTUAL MODULATION OF PERCEPTION ACROSS THE VISUAL FIELD: THE UNIFORMITY ILLUSION

Images in the visual system are not constructed like in a photographic camera, by an independent pointwise encoding of the input received in each receptive field. Rather, the presence of a stimulus in a region of the visual field can alter perception in a different location. For instance, in early visual areas, elements presented outside the classical receptive field of a neuron exert contextual modulatory effects on its neural responses

(Gheorghiu, Kingdom, & Petkov, 2014), by lateral connections and top-down feedback from downstream areas (Angelucci & Bressloff, 2006). Thus, response to an orientation signal by a sensitive neuron is suppressed by the presence of a contour with the same orientation (iso-orientation surround suppression) (Gheorghiu et al., 2014; H. Tanaka & Ohzawa, 2009). Peripheral crowding is another example of perceptual alteration induced by the concurrent presence of another stimulus: in the peripheral field, a form (for example, a letter) that would be easily identifiable in isolation becomes distorted if another form is presented in an adjacent location, while its apparent contrast and saliency remains the same (Petrov, Popple, & McKee, 2007; Whitney & Levi, 2011). In the phenomenon of perceptual filling-in, perception of a visual trait such as luminance, colour, motion or texture extends from a region of the visual field to a different region where it is not physically present (Komatsu, 2006). For example, maintaining stable fixation on an annulus surrounding a region of a different colour may result in the boundary between annulus and disc fading and the colour of the annulus spreading inwards and replacing the colour of the inner region. Filling-in is responsible for us not being aware of the existence of the blind spot, a region in the retina that lacks photoreceptors and therefore receives no visual input. It is also the cause of diverse illusions such as neon colour spreading (van Tuijl & Leeuwenberg, 1979), the Craik-O'Brien-Cornsweet effect (Grossberg & Todorovic, 1988) or the phantom illusion (Tynan & Sekular, 1975).

Many of these effects can be explained in terms of contour and surface processing (Gheorghiu et al., 2014). In real-life environments, scenes are composed of cluttered elements located at different depths, partially occluded by other objects. Consider, for example, a meadow showing behind a group of trees, with sunny areas alternating with shadows cast by the clouds above. Such a surface can appear discontinuous due to occlusions, and its low-level properties (luminance, colour) may vary greatly between regions: the darkest area of the meadow may have a level of brightness more similar to the trunks in the front than to the sunniest grassy area. And yet it is advantageous for survival that the brain is able to interpret the meadow as a unitary surface, even for

high-level operations, like predicting the trajectory of an animal running across. Progressive changes in luminance and colour must therefore be downplayed (if different regions are to be processed as a single surface). In relation to this, filling-in can occur in discontinuous patches of the same perceived surface, bridging over luminance gaps - such as those caused by occluding trunks (R. Kanai, Wu, Verstraten, & Shimojo, 2006). Conversely, it is necessary that our visual system interprets sharp gradients in luminance as physical borders (contour processing), and downplays the saliency of repeated orientation signals indicating the presence of a texture (iso-orientation surround suppression): for example, the blades of grass, or the stripes in the bark of the trunks must be suppressed in contrast to the contour of the trees (Gheorghiu et al., 2014; H. Tanaka & Ohzawa, 2009). Although crowding is ultimately a limitation in acuity due to pooling of adjacent signals, it has also been related to texture processing (L. Parkes et al., 2001).

Another imperative for contextual modulation (both in time and space) is related to feature binding: feature-dimensions such as colour, size, motion, etc, are processed separately in the early visual areas, yet they must be perceived together within an object. As Wu and colleagues have noted: the redness, roundness and motion of a rolling red ball are not perceived as separate entities (D. A. Wu, Kanai, & Shimojo, 2004). Aiming to investigate the mechanisms of feature-binding by exploring instances wherein it fails, these authors presented a striking example of steady-state misbinding of colour and motion across the visual field: a collection of red dots moving upwards in the central field and downwards in the periphery, together with green dots moving downwards in the centre and upwards in the periphery, was consistently seen as red-upwards and green-downwards across the whole visual scene. This illustrated that feature binding is influenced by contour and surface processing, so that binding takes place in a uniform way across regions assumed to be the same surface, even if this causes a non-veridical reassignment of features between objects. Furthermore, surface processing across the visual field is also affected by the different sensory precision in fovea and periphery: the illusion presented by Wu and colleagues suggests that recombination of features is

informed by the higher precision in the fovea along with an assumption of uniformity across the visual field (D. A. Wu et al., 2004).

This dovetails with another central aspect of our visual experience, whose basis is still poorly understood, namely fovea-periphery integration. Our subjective impression is that we see richness even in the most eccentric regions in the visual field, despite the fact that only the central 5 degrees (approximately) can perceive objects in detail. Although a partial explanation for the apparent richness of vision may be sequential acquisition of detail through exploratory saccades (O'Regan & Noe, 2001), an identical subjective experience can be obtained in absence of physical detail in the peripheral area of a scene: this is the case if an image is altered so that its peripheral traits are replaced by their summary statistics pooled over relatively large regions (Freeman & Simoncelli, 2011). Is the information in the peripheral regions of early sensory areas reconstructed based on the detail of the fovea, selectively applied according to the summary statistics of each region? This might entail a filling-in process in a discontinuous surface, as described in (R. Kanai, Wu, et al., 2006). Is this combination achieved in higher-level areas, where visual processing is non-local (or less local) (Ogmen & Herzog, 2010) and if so, is it fed back to early areas, or, rather, is it further processed without need for altering early sensory encoding? Or, alternatively, is our visual experience mostly a 'grand illusion' (Noë, 2002), wherein apparent richness is caused by decisional and metacognitive biases, not grounded on reconstruction of perceptual content but on perceptual inflation (Odegaard et al., 2018; Solovey et al., 2015)? Available evidence suggests that all of these processes may occur to some extent, but their specific contribution is debated (Michael A. Cohen et al., 2016).

A recently described visual illusion, named the Uniformity Illusion (UI), may help to shed light on this issue (Otten, Pinto, Paffen, Seth, & Kanai, 2016). When a presentation pattern has different but related properties in the fovea and periphery, after maintaining fixation for a time the appearance of the pattern changes so that the peripheral stimuli seem to take on the properties of the fovea. For example, a grid of

circles, with the peripheral circles being bigger than the central ones, will appear to change so that all circles are perceived as the same (smaller) size that is shown in the central area (see http://www.uniformillusion.com/ for other examples).

Although the mechanism of UI is still unknown, a collection of experiments described by Otten and colleagues (Otten et al., 2016) have revealed several interesting properties. First, UI occurs for many visual features, processed at low and higher-level sensory areas, including luminance, colour, size, shape, orientation, motion, text, etc. This broad variety suggests three possible explanations for the basis of UI: (i) it is either a widespread mechanism that operates independently at different brain areas, or (ii) a mid to high-level process taking place in downstream areas (common to very different types of visual processing), or, finally, (iii) a heterogeneous collection of processes that give rise to a superficially similar phenomenology.

Second, UI requires maintaining gaze fixation for a time, usually over 2 seconds, after onset of the presentation. This suggests that adaptation to peripheral input, with the consequent reduction of visibility of peripheral elements, may be a necessary step in the appearance of the illusion (Ditchburn & Ginsborg, 1952) -a process akin to Troxler fading (Balas & Sinha, 2007; Martinez-Conde & Macknik, 2017). Importantly, however, this does not lead to the disappearance of these peripheral elements, but to a change in their apparent properties, informed by the properties of foveal elements.

Third, the area over which appearance is altered is larger than that involved in similar illusions, like most instances of filling-in in retinally stabilized images.

Fourth, UI is enhanced (appears more frequently and with shorter delay) if the properties of central and peripheral patterns are more similar (for example, the size of the circles). Variance seems to play a role as well, especially variance of the foveal elements, so that low-variance presentation in the fovea (central elements of uniform

size, shape, orientation, etc) enhances UI. These properties suggest that visual variability, or precision, modulates perceptual uniformity.

Finally, according to metacognitive reports, illusory uniformity appeared similar, but (likely) not identical, to physical uniformity.

Otten and colleagues proposed a possible mechanism for UI, based on a predictive-coding account of perception (Otten et al., 2016). They suggested that adaptation deteriorates peripheral information faster than foveal input (Schieting & Spillmann, 1987), rendering the former weak and imprecise, so that bottom-up peripheral signals are overcome by a perceptual prediction based on the more reliable foveal elements (affected by less internal noise, and in most UI designs, also by less external noise, as the illusion is potentiated with a low-variance foveal pattern). This fovea-dominated prediction is extended to the periphery, likely in combination with a default prior for surface uniformity across the visual field.

However, this compelling explanation remains to be tested, and more precisely, it remains to be seen whether UI requires a reconstruction of perceptual content (and if so, at what level(s) of perceptual processing does this happen), or rather, if it results, in part or entirely, from a post-perceptual (decisional or meta-cognitive) bias in favour of uniformity across the visual field (perceptual inflation) (Odegaard et al., 2018; Solovey et al., 2015), wherein local variability is dismissed, not recoded. All of these questions are relevant for the basis of foveal-peripheral integration that is at the core of naturalistic visual experience.

4. GENERAL AIMS AND STRUCTURE OF THIS THESIS

As mentioned before, the general aim of this thesis is to explore contextual modulation of visual variability over time and across the visual field. Specifically, we focus on the two previously described instances of contextual modulation: serial dependence and UI.

Part II presents a series of experiments examining serial dependence in the perception of visual variance: in other words, how perceptual decisions about visual ensemble variance are influenced by previous variance presentations. We hypothesize that a higher-order statistic such as variance may be subject to history-dependent biases similar to those observed for other visual dimensions, in order to construct a flexible yet largely stable experience of environmental variability out of sequential pieces of statistical evidence. Furthermore, we aim to examine the specific levels of perceptual processing that generate serial dependence in variance (and serial dependence in general) and the influence of both stimulus properties and high-level processes related to metacognition and perceptual awareness.

Part III presents a study on the mechanistic basis of UI, whereby visual variability across the visual field is apparently overwritten by uniform, precise foveal information. We employ an adaptation paradigm in order to explore whether UI on orientation (a V1-based feature-dimension) and texture-density (dimension processed beyond V1) is associated to local changes in sensory encoding in peripheral receptive fields.

Although not part of the core of the thesis, we also conducted an additional research project exploring how visual variability (rate of perceptual change over time and across the visual field), drives, in turn, the emergence of a high-level trait, central to conscious experience, such as the subjective impression of passage of time (perceived duration). In this study, we contrast the influence of (external) visual variability with that of fluctuations in physiological and neural processes that previous studies have claimed to modulate perceived duration. This additional research is detailed in Part V.

PART II: Contextual Modulation Over Time: Serial Dependence in the Perceptual Processing of Visual Variance

Abstract

In Part II we examine contextual modulation of visual variability over time: in other words, how variance perception is dependent on previous variance presentations (serial dependence). Several experiments reported here have been published in a Journal of Vision article (2018) with the title 'Serial Dependence in Visual Variance' (Suárez-Pinilla, Seth, & Roseboom, 2018b). By requiring judgments about the motion variance of sequentially presented random dot kinematograms (RDKs), we found two different history-driven biases operating at different timescales: a positive (attractive, Bayesian-like) bias toward very recent variance presentations and a negative (repulsive) effect driven by the broader, less recent context. The positive bias (corresponding to the serial dependence effect reported for other visual dimensions) was independent of low-level stimulus properties, arose only in relation to high-confidence decision-making and seemed subject to time and capacity limitations, suggesting a high-level origin (likely decisional representations interacting in short-term memory). The negative bias may be consistent with sensory adaptation. In a continuous flash suppression study about serial dependence in orientation, we further confirmed that, unlike adaptation after-effects (which persist for masked stimuli), serial dependence did not appear in relation to unseen trials, indicating that it cannot arise merely as a result of local low-level sensory changes. We propose that the negative and positive serial dependencies arise at different levels of perceptual decision-making (sensory adaptation and Bayesian-like decision, respectively) to handle the necessary balance between sensitivity to change and perceptual stability. A two-layer model based on this assumption was able to reproduce the observed biases in their respective timescales, utilising population codes to produce a sensory response and iterative Bayesian operations for generating a decision based on such sensory information.

CHAPTER 1: SERIAL DEPENDENCIES IN THE PERCEPTUAL PROCESSING OF VISUAL VARIANCE

Variance is a key statistic to the interpretation of sensory information, yet the mechanisms whereby the visual system processes this property are still poorly understood. We aimed to gain insight on variance perception as a distinct visual feature by studying how variance judgments are influenced by previous variance stimuli (serial dependence). In this chapter we report the first of a series of experiments (Experiment 1), wherein participants were asked to estimate the variance of motion trajectories of sequentially presented random dot kinematograms. Two opposite history-driven biases were observed. A positive (attractive) bias was exerted by very recent presentations, similar to other instances of serial dependence; this effect was apparently tuned by perceptual similarity within the same dimension (i.e. variance), but independent on low-level or associated stimulus properties (eccentricity, spatial location, similarity between means of consecutive stimuli). A negative (repulsive) bias was observed in relation to less recent presentations. A series of control analyses revealed that both the positive and negative dependencies arose only in the true trial history (and not in relation to future or shuffled trials) and were not dependent on stimuli being sampled from a closed set. The majority of the contents of this chapter have been published in the Journal of Vision as part of the article 'Serial Dependence in the Perception of Visual Variance' (Suárez-Pinilla et al., 2018b).

1. INTRODUCTION

Considerable evidence indicates that the human visual system is able to extract statistical information from sensory signals supporting the formation of summary or 'ensemble' representations across a variety of dimensions, including low-level features such as orientation or size, as well as higher-level (complex or abstract) traits such as facial expressions in a crowd (George A. Alvarez, 2011; George A. Alvarez & Oliva, 2009; Ariely, 2001; S. C. Chong & Treisman, 2003; Geisler, 2008; Rosenholtz et al., 2012). Such information can be used to efficiently encode (Dahmen et al., 2010; Fairhall et al., 2001) and interpret subsequent sensory inputs, and to make predictions about future events (Summerfield & Egner, 2009; Summerfield & Lange, 2014).

Many forms of visual input can be summarised in terms of statistical moments such as central tendency (e.g., mean) and variance or dispersion (consider, for example, a random dot kinematogram, which will have a mean and a variance in the distribution of dot motion). Most studies on ensemble processing have focused on central tendency statistics (Albretcht & Scholl, 2010; S. C. Chong & Treisman, 2005b; Jennifer E. Corbett & Oriet, 2011; Haberman & Whitney, 2009; Im & Chong, 2014; Sweeny & Whitney, 2014; Wolfe, Kosovicheva, Leib, Wood, & Whitney, 2015), while variance computations have received less attention. However, variance is known to play a crucial role in visual experience, modulating perceptual grouping, ensemble averaging (Timothy F. Brady & Alvarez, 2015; Gardelle & Mammasian, 2015; Gardelle & Summerfield, 2011; John Maule & Franklin, 2015; J Maule et al., 2014; Zylberberg et al., 2014), texture processing (Michael J. Morgan et al., 2014; M.J. Morgan et al., 2012) and comparison between arrays (Fouriezos et al., 2008). Variance is also critical to perceptual prediction, since it provides a measure of the expected range of stimuli (Summerfield & Lange, 2014) as well as the precision (reliability) of the sensory input (Jennifer E. Corbett et al., 2012; Meyniel et al., 2015; Sato & Kording, 2014). As an indication of sensory reliability, variance also affects metacognitive judgments, although evidence is conflicting

regarding the extent and direction of this effect (Gardelle & Mammasian, 2015; Spence et al., 2016; Zylberberg et al., 2014).

Notably, most studies involving variance manipulations have examined its impact on perceptual decisions about other features, rather than on the perception of variance itself. Comparatively few studies have investigated the mechanisms of variance processing directly. Those that do have addressed mainly three questions: what are the general properties of variance processing (speed, automaticity, attentional demands), to what extent does variance computation rely on the processing of the individual elements of the ensemble, and whether it operates as a specific trait of the ensemble or the feature-dimension over which it is computed, or rather as an abstract property. So far, these studies have employed heterogeneous designs and reached disparate conclusions. Concerning the general properties of variance processing, a study examining judgments of colour diversity (Bronfman, Brezis, Jacobson, & Usher, 2014) suggested a rapid, pre-attentive mechanism. This is in agreement with another study which reported priming by visual variance, an effect that seems to occur rapidly, automatically and without need of feature-based attention (Michael et al., 2014); however, this latter study did not examine variance perception directly, but only the priming effect of variance on mean judgments. Regarding the second question, namely the reliance of variance computation on the processing of individual elements, available evidence (based on highly heterogeneous studies) is conflicting: one study on pattern regularity (positional variance) suggested a very inefficient computation of variance, underwritten by subsampling of a small fraction of the elements of the array (M.J. Morgan et al., 2012). By contrast, and surprisingly given the finding of a rapid, pre-attentive mechanism, the abovementioned study on colour diversity reported that variance processing required a conscious representation of the individual elements (Bronfman et al., 2014); in similar manner, a study on facial emotions in a crowd determined that variance judgments along this dimension relied on high-level processing of individual faces (Haberman, Lee, et al., 2015). Finally, regarding the question about whether variance, once computed over a certain feature-dimension of

a visual ensemble, retains its specificity or rather emerges as an abstract property, the previously reported study on priming suggested that the effect of (implicit) variance on mean perception was feature-specific (Michael et al., 2014). In contrast, a study on direct variance perception found negative adaptation after-effects which generalized across dimensions of visual variance, suggesting a high-level, rather than sensory origin for this effect, and therefore, that variance operates as an independent cognitive property (Payzan-LeNestour et al., 2016). In summary, available evidence shows some dissension, but a picture starts to emerge: variance computations would be relatively rapid, but appear to require high-level processing of the individual elements of the array; however, once computed, variance would become dissociated from the properties of the ensemble and of the perceptual dimension over which it was estimated, and operate as a high-level cognitive trait.

To clarify some the above issues, here, we examine variance processing as a distinct perceptual feature, by investigating the influence of previous variance presentations on judgments about this dimension. It has long been known that perception is affected by previous input. Influences of past perceptual events on current perception fall generally into two different categories. Most well-known are adaptation after-effects - repulsive (negative) biases in perception exerted after prolonged exposure to a certain stimulus magnitude - that have been described for many low and high-level traits (including variance) (Campbell & Maffei, 1971; Mather, Verstraten, & Anstis, 1998; Payzan-LeNestour et al., 2016; Roseboom, Linares, & Nishida, 2015) and which are classically employed as an experimental tool for investigating perceptual mechanisms (Kohn, 2007). The second category is observed in relation to shorter presentations, generally consisting of an attractive (positive) perceptual bias toward recent sensory history. These ‘serial dependencies’ have been found for several low and high-level features (Cicchini et al., 2014; Fischer & Whitney, 2014; John-Saaltink et al., 2016; Alina Liberman et al., 2014; Xia et al., 2015). It has been proposed that these two different effects contribute in opposite ways to the ‘tuning’ of the balance between perceptual sensitivity and stability: while negative adaptation produces a normalization of neural

representations in order to maximize sensitivity to changes around the most frequent stimulus intensity, serial dependence contributes to perceptual stability by smoothing out discrete discontinuities as sensory noise (Fischer & Whitney, 2014). However, the specific level(s) of perceptual processing at which serial dependencies arise (low-level perception, attention, working memory, decision, response) are still unclear and a matter of intense debate (Bliss et al., 2017; Cicchini et al., 2017; Fischer & Whitney, 2014; Fritsche et al., 2017; John-Saaltink et al., 2016; A. Kiyonaga et al., 2017; Manassi et al., 2018).

Our study employs serial dependence in variance judgments as a way to track the dynamics and timescales of the processing of this statistic as a distinct perceptual feature. In the current chapter we report the results for Experiment 1, which investigated the existence, magnitude and direction of serial dependence in visual variance perception (operationalized as variance in RDK motion direction), as well as its relationship with associated stimulus features, such as ensemble mean and spatial location. For this experiment we separately explored fovea and periphery, as the compression of visual information into summary statistics is particularly relevant to the much larger, poorly spatially-resolved peripheral field (Balas et al., 2010; Freeman & Simoncelli, 2011; Rosenholtz et al., 2012; C. Ziemba & Simoncelli, 2015). In subsequent chapters (2-4) we describe further experiments aiming to elucidate the specific perceptual mechanisms that give rise to serial dependence and the variables that affect the maintenance of the effect in time. Finally, in Chapter 5 we present a model based on our premises about the origin of different history-dependent biases, which reasonably captures the magnitude and timescales of the previously described serial dependencies in variance.

2. EXPERIMENT 1: SERIAL DEPENDENCIES IN VARIANCE JUDGMENTS

In this experiment we investigated the existence of serial dependence in variance judgments and its relationship to basic features of stimulus presentation, including eccentricity, spatial location, and ensemble mean. We operationalized visual variance as variance in the motion direction of a random dot kinematogram (RDK), by manipulating independently means and variances on a trial basis and asking participants to report the 'randomness' of the presented motion -see Figure 1 for an overview of the trial structure.

2.1. METHODS

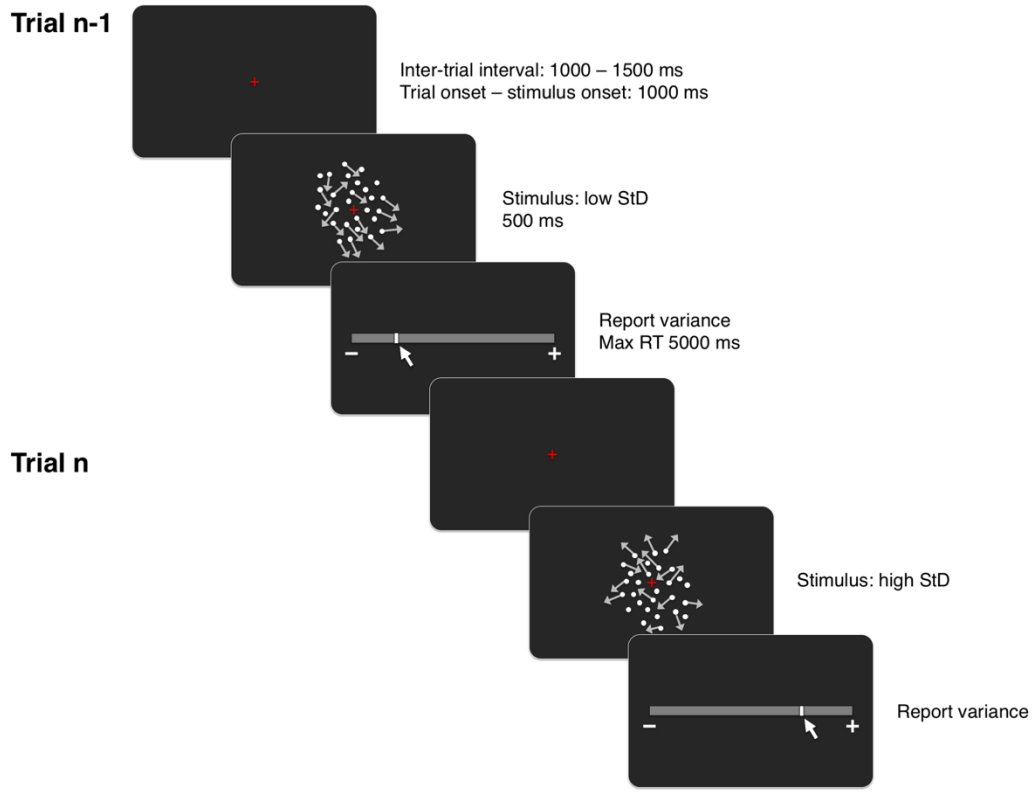


Figure 1. Experiment 1: Methods. Each trial presented an RDK of a certain mean and variance (standard deviation, StD) in the motion trajectories of its component dots. In the example, trial n-1 and n have low and high StD values, respectively. The stimulus was presented for 500 ms. In half of the blocks, the RDK was displayed in the fovea (0 degrees of visual angle, dva) whereas in the other half it was displayed in the periphery (20 dva, equally frequent in the left and right hemifield). Experiment 1 required variance (StD) reports for each trial, using a visual analogue scale. Abbreviations: StD – standard deviation (of the RDK motion directions), RT: response time, ms: milliseconds.

2.1.1. Stimuli

The stimulus consisted of a cluster of random moving dots (RDK) displayed for 500 milliseconds at a certain eccentricity (0^o or 20^o, see below) over a dark grey background

(3.92 cd/m^2). The cluster spanned 5 degrees of visual angle (dva) along the horizontal and vertical dimensions and was comprised of 100 light grey dots (diameter of 0.11 dva, luminance 43.14 cd/m^2) moving along a straight trajectory at a rate of 2 pixels by frame (8.45 dva/s). The initial position of each dot was uniformly randomized (excluding overlap with other dots) and its coordinates were updated per frame by a trigonometric calculation based on the individual dot's angular motion direction, re-entering the cluster from the opposite side if it reached a boundary. Each dot's motion direction was extracted from a circular Gaussian (von Mises) distribution that varied for each stimulus presentation: its mean could take any random integer value from 0º to 359º and its standard deviation was pseudorandomized among 6 possible values, namely 5º, 10º, 20º, 30º, 40º and 60º. This parameter (henceforth **StD**) is the dimension of interest in this experiment.

2.1.2. Procedure

The experimental session comprised a practice block with 72 trials and eight experimental blocks with 60 trials each. The practice had a double purpose: (i) familiarizing participants with the scoring process and the scale used in the experiment, and (ii) training maintenance of centred gaze fixation. A broader range of StD values was presented during the practice block compared to the experimental blocks (1º, 10º, 20º, 36º, 60º and 90º). In both practice and experimental blocks, participants had to score the 'randomness' (variance) of the motion of the RDK using a visual analogue scale (see Figure 1), by adjusting the position of a sliding bar with the mouse. During the practice block, feedback was provided by showing the correct response (score corresponding to the veridical variance) on an additional scale which appeared below the one employed by the participant, after they had produced a response. For simplicity, the scale was a linear translation of the StD numeric values ranging from 0º (left end) to 90º (right end of the scale). Within this block, the first 36 trials were 'foveal' (the stimulus was presented at 0 dva eccentricity) and the remaining 36 were 'peripheral' (20 dva).

Regarding the eight experimental blocks, they employed the narrower set of StD values detailed in the previous 'Stimuli' section (5°, 10°, 20°, 30°, 40° and 60°) and did not have feedback. Half of the eight blocks were 'foveal' (stimulus presentation at 0 dva eccentricity for all trials) and half 'peripheral', with presentation at 20 dva along the horizontal axis, equally frequent in the right and left hemifields. The sequence of foveal and peripheral blocks was pseudorandomized for each participant.

Eye tracking was performed during the entire experimental session. Calibration of the eye-tracking system was performed at the beginning of each block (practice and experimental) using a standard 5-point grid, allowing for a maximal average error of 0.5 dva.

At the beginning of each trial a red fixation cross appeared on the centre of the screen, spanning 1.1 dva (horizontally and vertically). Participants were instructed to maintain their gaze on the fixation cross. The RDK stimulus appeared after 1000 ms, and both the stimulus and the fixation cross disappeared simultaneously at 1500 ms from trial onset. Immediately after, the response scale and sliding bar were displayed on the screen. The initial position of the bar was randomized for each trial along the whole length of the scale to exclude the possibility that participants simply repeated the same (response) action on each trial. If the participant failed to respond within 5 seconds, the next trial started automatically. The inter-trial interval was randomized between 250-1000ms.

On each trial, participants were allowed to correct their gaze position during the first 700 ms, if they noticed that their gaze had deviated from the central fixation cross. However, if a deviation (of more than 5 dva) occurred between 700 - 1000 ms, the trial was aborted and restarted. About a third of participants (9/30) were tested with a slightly different procedure, in which a trial abortion led directly to the start of the next trial (after the inter-trial interval). This procedure led to the exclusion of more trials from analysis, since poorly fixated trials were not restarted. Importantly, in both cases trials

retained for analysis were those in which fixation was maintained during stimulus presentation (1000 ms – 1500 ms), and no trial was aborted or restarted after stimulus onset at 1000 ms.

2.1.3. Participants

Participants were recruited through online advertisement and among members of the laboratory. All were over 18 years old and reported normal or corrected-to-normal vision. Every participant signed an informed consent form before taking part and was either awarded 10 course credits or paid £10 for their participation. The study was granted ethical approval by the Research Ethics Committee of the University of Sussex.

2.1.4. Apparatus

Experiments were programmed in MATLAB 2012b (MathWorks Inc., Natick, US-MA) with Psychtoolbox 3.0.10 and displayed on a LaCie Electron 22BLUE II 22'' with screen resolution of 1024 x 768 pixels and refresh rate of 60 Hz. Eye tracking was performed with Eyelink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at a sampling rate of 1000 Hz, using a level desktop camera mount. Head position was stabilized at 43 cm from the screen with a chin and forehead rest.

2.1.5. Statistical analysis

Statistical analyses (detailed in the Results section) were performed on Matlab 2016a (MathWorks Inc., Natick, US-MA), R 3.4.2 (The R Foundation for Statistical Computing, http://www.R-project.org) and JASP (JASP Team (2017). JASP (Version 0.8.3.1, Mac OS X – El Capitan (10.11)).

2.2. RESULTS

Thirty participants (25 female, mean age 19.0 y/o, standard deviation 1.35) participated in this experiment. Except for two members of the laboratory, the rest were first-year Psychology students who volunteered for course credits.

To ensure the validity of foveal and peripheral conditions, trials without centred gaze fixation during stimulus presentation were removed from the analysis: a trial was deemed valid if the participant maintained fixation within 5 dva from the centre of the screen for over 80% of the stimulus presentation period (1000-1500 ms from trial onset). Invalid trials were removed, as well as all data regarding trial history that involved at least one of these trials: for instance, if trial n was valid but trial n-3 was not, trial n was not included in analyses regarding serial dependence associated to position n-3 or further backwards. A total of 12480 trials entered the analysis.

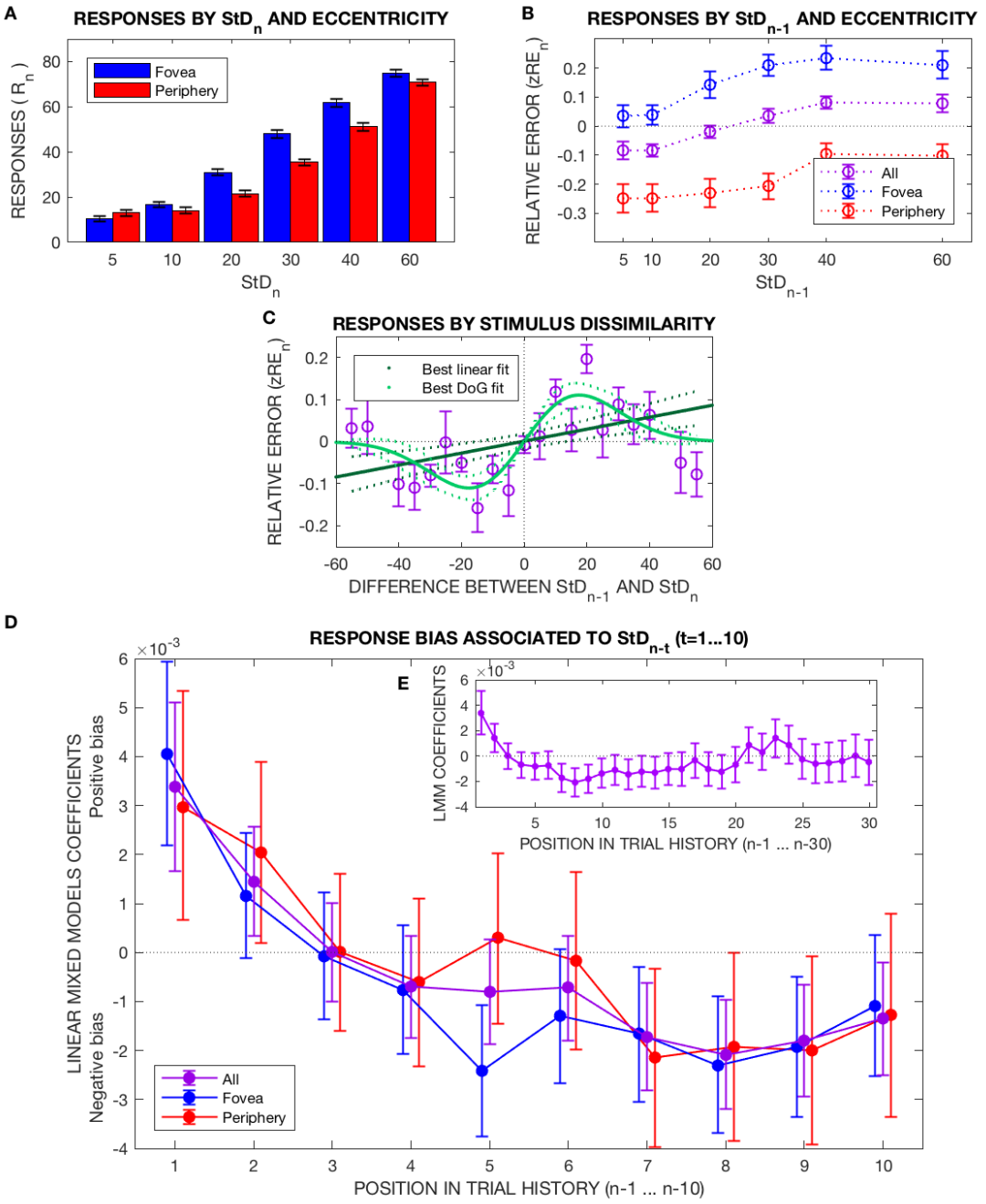


Figure 2. Experiment 1: Results. 2a. Distribution of responses by StDn and eccentricity. The height of the bars represents the mean and the error bars the between-subject standard error. **2b.** Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}). The relative error, defined as $RE_n=(R_n-StD_n)/StD_n$, has been normalized by the distribution of errors provided by each subject for the current StD_n; thus, a positive zRE_n means a larger report in that trial than the participant's average for that stimulus level, and conversely a negative zRE_n indicates a lower-than-average score i.e., sign is not necessarily related to comparison with veridical StD_n, if the participant exhibits a systematic bias for that StD_n. Consequently, plotting zRE_n reports by StD_{n-1} allows examination of any possible bias in relation to previous trial StD_{n-1}, beyond any unrelated source of bias. The error

bars represent the between-subject standard error. The ascending slope of the plots indicates a positive bias associated with $StD_{n-1,}$ for both foveal and peripheral presentations: relative overestimation occurs for larger StD_{n-1}. **2c.** Normalized response error in the current trial (zRE_n), plotted as a function of the difference between previous and current StD: $\Delta StD_{n-1,n}=StD_{n-1}-StD_n$. Visual inspection of the data shows that, when the previous RDK had a larger variance, judgments about the current variance were overestimated, and vice versa. In other words, the sign of zRE_n follows that of $\Delta StD_{n-1,n}$, again confirming a positive bias driven by the previous trial. However, such effect is likely non-linear, as the peak amplitude of the bias occurs at intermediate values of $\Delta StD_{n-1,n}$, both positive and negative. Conversely, the bias is dampened (close to zero) when previous and current StD are highly dissimilar. In order to better characterize the shape of the serial dependence effect, the graph shows the best-fitting linear and derivative of Gaussian (DoG) functions for the data. In terms of R^2, the DoG equation shows a better fit, as reported in the main text: this suggests that serial dependence is non-linearly tuned by StD (dis)similarity. Nevertheless, these results must be taken with caution due to confounding by floor and ceiling effects in responses to extreme StD values (see main text for details). **2d.** Response bias associated with StD presented in recent history. Each data point represents the fixed-effects coefficient estimate (B) in a Bayesian linear mixed-effects model (LMM) for the association between the StD presented in trials n-1 to n-10 (StD_{n-t}, t=1…10) and the normalized response error in the current trial. The value of the B coefficient represents the linear slope between the past StD at certain trial position (StD_{n-t}) and the normalized response error provided in the current trial: i.e., the variation (in z-scores) observed on the current response (regardless of the presented StD), when StD_{n-t} was increased by 1^o. A positive B represents an attractive bias (ascending slope), and a negative B a repulsive bias (descending slope). The error bars depict the 95% credible intervals for the value of the B coefficient. The inset graph (**2e**) presents the Bayesian LMM coefficients and 95% credible intervals for the association between StD_{n-t} and current zRE_n, up to trial n-30, foveal and peripheral data pooled.

2.2.1. Overview of responses

Figure 2a shows the distribution of responses (R_n) for each StD value and visual eccentricity. Showing that participants were able to perceive the different levels of variance presented in the experiment, reports were positively correlated and monotonically increased with stimulus StD for both foveal and peripheral presentations.

To examine the general pattern of variance judgments, we conducted a repeated-measures ANOVA on the influence of two within-subject factors, StD in the current trial (StD_n) and eccentricity, on participant's responses. Both main effects and their interaction were significant (sphericity correction was applied by the method of

Greenhouse-Geisser). For StD_n, the main effect yielded $F(1.825,45.621)=473.80$, $p<0.001$, $\eta^2_p=0.950$, in relation with higher reports for larger stimulus StD. For eccentricity the main effect was $F(1,25)=33.32$, $p<0.001$, $\eta^2_p=0.571$: peripheral presentation was associated with lower variance reports, with a mean difference of 6.798 (fovea - periphery), $t(25)=8.237$, $p<0.001$, Cohen's $d=1.615$. The interaction effect StD_n*eccentricity was also significant, $F(2.715, 67.882)=20.06$, $p<0.001$, $\eta^2_p=0.445$, indicating that the difference between foveal and peripheral responses increased for large StD_n values, as shown in figure 2a. These results were confirmed in a Bayesian RM ANOVA with the same variables: the full model (both main effects and interaction) was the most explanatory according to the Bayes factor, outperforming the second best (only the two main effects) by a factor of $BF_{full/main\ effects}=1.075*10^6$. These findings (lower responses in periphery than in fovea, especially for large StD_n) seem to relate to a greater regression to the mean exhibited in responses about peripheral stimuli (likely due to worse discrimination between stimulus levels), combined with the fact that the range of the response scale allows for larger errors by overestimation than underestimation.

To further characterize perception of variance throughout the different presented StD levels and confirm the apparently worse performance in the periphery, we examined the dispersion of the responses per StD (σ_R), defined as the standard deviation of the distribution of responses per stimulus level. We conducted a repeated-measures ANOVA on the influence of StD_n level and eccentricity on response dispersion. The main effect for StD_n yielded a $F(2.994,74.840)=58.426$, $p<0.001$, $\eta^2_p=0.700$, in relation with greater response dispersion for large StD levels, as is common in magnitude estimation tasks: σ_R was lowest at 9.87 for StD=5 and steadily increased with StD value until it peaked at 21.03 for StD=30, remained almost equal (σ_R= 20.51) for StD=40 and decreased moderately for StD=60 (σ_R= 15.98), probably due to a ceiling effect. As for the main effect of eccentricity on response dispersion, it was $F(1,25)=4.165$, $p=0.052$, $\eta^2_p=0.143$, suggesting a trend towards greater response dispersion in peripheral presentations: mean difference -0.658 (fovea – periphery), $t(25)=-1.738$, $p=0.086$,

Cohen's *d*=-0.339. Last, the effect of the interaction term StD_n* eccentricity was *F*(3.530, 88.244)=4.757, *p*=0.002, η^2_p=0.160, due to the larger response dispersion in periphery occurring mainly for large StD values. In a Bayesian RM ANOVA with the same variables as in the frequentist counterpart, the best model was the full model (StD_n, eccentricity and interaction), which outperformed the second best (with only StD_n) by a factor of $BF_{full/StDn}$=8.747. In summary, response dispersion increased with stimulus (StD) level and there was a (nearly significant) trend towards greater response dispersion for peripheral presentations, especially at large StD, suggesting a slightly worse performance at 20 dva eccentricity compared to 0 dva, in agreement with the previous finding of a greater regression to the mean in peripheral responses.

2.2.2. Variance reports are subject to a positive bias driven by very recent trial history

To characterize the existence of serial dependences in variance reports, we tracked whether the response errors provided by each participant for each StD level were different as a function of the StD level presented in the previous trial (serial dependence in relation to trial n-1), or in positions further backwards in trial history (trial n-t). Unless stated otherwise, the response variable in our analyses of serial dependence is the normalized response error relative to the current stimulus (zRE_n). Response errors (defined as $RE_n=(R_n-StD_n)/StD_n$) are normalized by the distribution of reports provided by each individual for the level of StD presented in the current trial. Thus, zRE_n sums to zero across all trials for a given participant and StD_n level: a negative zRE_n indicates that the participant provided a below-average response in that trial compared to their responses for other physically identical stimuli, while a positive zRE_n indicates an above average response. Therefore, normalization ensures that the value of the response variable, zRE_n, is independent of the current StD_n level and of each participant's global scoring biases.

Figure 2b presents the average zRE_n as a function of the previous stimulus (StD_{n-1}), plotted separately by eccentricity. Regardless of generally lower reports at larger eccentricity, a trend towards larger zRE_n for higher StD_{n-1} values is evident for all trials pooled as well as for both foveal and peripheral presentations, as shown by the ascending slope of the three plots (Fovea, Periphery, All). In other words, there was a relative overestimation of the current stimulus when the previous stimulus had a large StD, and a relative underestimation when the previous StD was small, compared to other trials in identical conditions of eccentricity. This indicates a positive (attractive, Bayesian-like) bias driven by trial n-1: current responses resemble the previous stimulus – serial dependence for visual variance.

To verify this observation, we ran a repeated-measures ANOVA on the effect of StD_{n-1} level (as within-subject factor) on current variance reports (zRE_n). The effect of StD_{n-1} was statistically significant ($F(3.231,93.697)=7.221$, $p<0.001$, $\eta^2_p=0.199$, Greenhouse-Geisser correction applied). The Bayes factor for the inclusion of StD_{n-1} compared to the null model (both of them included participant as grouping variable) was $BF_{inclusion}=56187.91$, indicating *extreme* (Wagenmakers et al., 2017) evidence for a superior explanatory ability of the model that included this term.

Serial dependence in variance may be non-linearly tuned by stimulus (dis)similarity

Our analyses so far have shown that variance judgments are attracted to stimuli presented in the previous (n-1) trial, demonstrating the existence of serial dependence in variance as has been found for several visual dimensions. Other studies have observed that this attractive effect is non-linearly tuned by the difference between previous and current stimuli, so that the size of the bias is maximal when both are relatively similar, and is reduced or even disappears if the difference between magnitudes is very large (Cicchini et al., 2014; Cicchini et al., 2017; Fischer & Whitney, 2014; Fritsche et al., 2017).

We enquired whether such dependency of stimulus (dis)similarity was also observed in our experiment by analysing normalized response errors in the current trial (zRE_n) as a function of the difference between previous and current stimuli: $\Delta StD_{n-1,n} = StD_{n-1} - StD_n$. A Bayesian RM ANOVA on the effect of $\Delta StD_{n-1,n}$ (as within-participant factor) on zRE_n showed extreme evidence compared to the null model (BF_{10}=333.055), unsurprising since we had already found an effect of StD_{n-1}, and zRE_n reports are normalized to be independent of StD_n. This result by itself does not tell us anything about the shape of the relationship between both magnitudes. Figure 2c graphically depicts this shape, presenting the average zRE_n (and between-participant standard error) as a function of $\Delta StD_{n-1,n}$. A positive $\Delta StD_{n-1,n}$ indicates that the previous trial RDK had a larger StD than the current one, and vice versa. Therefore, an attractive effect driven by the previous stimulus would produce zRE_n reports of the same sign as $\Delta StD_{n-1,n}$: i.e. relative overestimation of current variance when the previous variance was larger, and vice versa. However, if this attraction is non-linear, the effect may peak at non-extreme values of $\Delta StD_{n-1,n}$ and return to baseline for larger differences. A visual inspection of our results certainly suggests this pattern.

Previous studies have successfully modelled the non-linear relationship between stimulus dissimilarity and serial dependence by using a derivative of Gaussian (DoG) function (Fischer & Whitney, 2014; Fritsche et al., 2017). We contrasted the performance of a linear function,

(1)

$$y = ax + b$$

with a derivative of Gaussian function,

(2)

$$y = \frac{ax}{\sqrt{2\pi}\sigma^3} e^{\frac{-x^2}{2\sigma^2}}$$

by fitting both to our data. In both cases, *x* is the difference between previous and current StD ($\Delta StD_{n-1,n}$), and *y* is the current normalized response error (zRE_n). The coefficients to be fitted in the linear equation are the slope and intercept, *a* and *b*; in the DoG equation, *a* and σ are the amplitude and the width (standard deviation) of the function, respectively. Fitting was achieved by non-linear least-squares method.

Figure 2c shows the best-fitting linear (dark green) and DoG functions (light green), along with the 95% prediction intervals for the value of the function, computed non-simultaneously at each $\Delta StD_{n-1,n}$. Concerning the linear function, its slope was *a*=0.0014, with 95% confidence interval (0.0007, 0.0021), indicating a statistically significant attractive effect, represented by a linear increase of 0.0014 z-scores for each 1º of stimulus difference between previous and current StD. As for the DoG function, its coefficient estimates and 95% confidence intervals were *a*=141.5 (75.19, 207.8), σ=17.61 (13.43, 21.8). In other words, according to the derivative of Gaussian fit the maximum effect is exerted when the stimulus difference is 17.61º, entailing a peak attraction of 0.1104 (z-scores) -the direction of the bias is given by the positive sign of *a*. Comparing the adjusted R^2 of both models, the DoG model showed a better fit: $R^2_{adjusted}$=0.0054, versus $R^2_{adjusted}$=0.0014 for the linear equation. Thus, our results show a non-linear dependence of stimulus dissimilarity, similar to that observed in previous studies of serial dependence in other visual dimensions.

However, this result must be taken with caution, as reports in our task are subject to noticeable floor and ceiling effects. Since the most extreme values of $\Delta StD_{n-1,n}$ can only occur if the current StD_n is either very low or very high, the fact that such extreme values are associated to almost no response bias (as seen in Figure 2c) may be simply due to artificial constraints by the limits of the response scale. In fact, visual inspection of the data in Figure 2c shows that, whenever StD_n took the lowest value 5º (corresponding to $\Delta StD_{n-1,n}$ values of 5º, 15º, 25º, 35º and 55º), the response bias was close to zero. Thus, it is unclear to what extent the observed non-linearity may be explained simply by constraints in the response scale.

In conclusion, it is likely that serial dependence in variance is non-linearly tuned by stimulus dissimilarity, as observed for other visual dimensions; however, this result must be taken with caution. Henceforth we will continue to base our serial dependence analyses on the effect of StD_{n-1}, rather than $\Delta StD_{n-1,n}$, as we consider the first approach more parsimonious given the particularities of our task.

Serial dependence in variance does not depend on other stimulus properties (visual eccentricity, spatial location or ensemble mean)

Having established the existence of a positive serial dependence exerted by the previous trial, we sought to ascertain which properties of the stimulus presentation might modulate such bias. Previous studies on serial dependence have observed that it appears in the fovea as well as the periphery, and its strength is tuned by spatiotemporal proximity (Fischer & Whitney, 2014). Moreover, if the function of (positive) serial dependence is to promote perceptual continuity (Fischer & Whitney, 2014), it seems reasonable to expect that similarity of other attributes of consecutive stimuli would lead to a stronger influence of the studied feature-dimension, especially for two attributes as closely related as ensemble mean and variance.

To test the influence of these properties, we conducted a repeated-measures ANOVA on zRE_n (as dependent variable) with two within-subject factors: StD_{n-1} and each of the features of interest, separately: eccentricity, retinal location and similarity of means.

For eccentricity, both main effects were statistically significant ($F_{eccentricity}(1,25)=31.004$, $p<0.001$, $\eta^2_{p\ eccentricity}=0.554$; $F_{StDn-1}(2.662, 66.556)=7.029$, $p<0.001$, $\eta^2_{p\ StDn-1}=0.219$; sphericity, correction by Greenhouse-Geisser), while the interaction was not ($F(3.789, 94.722)=1.710$, $p=0.157$, $\eta^2_p=0.064$). This result, as indicated by the roughly parallel plots for fovea and periphery in figure 2b, suggests that while eccentricity influences the

absolute value of the current StD response, it does not modulate the serial dependence exerted by StD_{n-1}. To formally test this hypothesis, we turned to Bayesian repeated-measures ANOVA. Table 1a summarizes the comparison between all competing models. The largest Bayes factor corresponds to the model including both main effects but not the interaction ($BF_{10}=3.432*10^{29}$), which outperforms the model that also includes the interaction term StD_{n-1}*eccentricity by a factor of $BF_{main\ effects/full}$ =17.645 – strong (Wagenmakers et al., 2017) evidence against its inclusion and supporting the conclusion that while there is an overall difference in reports, there is no difference in serial dependence across eccentricity.

Regarding the influence of spatial location, we analysed only peripheral presentation blocks, classifying trials according to whether the previous stimulus had been presented on the same or the opposite hemifield as the current one: i.e. same presentation location versus a separation of 40 dva between consecutive presentations. Results for the model comparison given by a Bayesian repeated-measures ANOVA are presented in Table 1b: the best model in terms of evidence includes only StD_{n-1} ($BF_{10}=2.073$), while the worst model also includes the hemifield and the interaction term StD_{n-1*} hemifield ($BF_{10}=0.120$). This indicates moderate evidence against the full model (including interaction) compared to the null, and strong evidence against it when compared to the most explanatory model, i.e. the one with StD_{n-1} only ($BF_{full/StDn-1}=0.058$). These results support the hypothesis of serial dependence being unaffected by the spatial location of consecutive stimuli.

Last, we examined the influence of mean RDK direction on serial dependence of variance; specifically, whether the magnitude of the serial dependence effect (in variance) depended on the successive presentations containing a similar mean direction. With this aim, we binned the absolute difference between the mean RDK directions in the previous and current trial into 5 categories: ≤36º, 37º-72º, 73º-108º, 109º-144º, 145º-180º. As before, we conducted a Bayesian repeated-measures ANOVA with two within-subject factors (StD_{n-1} and mean difference). As shown in Table 1c, the

best model included only StD_{n-1} (BF_{10}= 3.210*10^5), whereas the model including both main effects and its interaction was the second worst after the one with mean difference only, with a BF_{10}=0.281. The Bayes factor for inclusion of the interaction term indicated extreme evidence against it ($BF_{inclusion}$=3.491*10^{-6}); this was also the case if the comparison was made between the full model and the model lacking only the interaction ($BF_{main\ effects/full}$ = 1789.55). This lack of association of mean similarity and serial dependence in variance was further confirmed in a supplementary experiment (named Experiment 1B and detailed in section 3.3 of the current chapter -see below) that used a limited range of mean trajectories, allowing for only four between-trial differences (0º, 35º, 55º and 90º).

In summary, serial dependence in variance reports is not modulated by low-level properties of the stimulus including visual eccentricity or spatial location, or associated features such as mean, suggesting that visual variance (operationalized as variance of motion direction) is processed as a feature dimension independent from these properties, at least at the level of perceptual decision-making that gives rise to serial dependence.

TABLE 1. Serial dependence and stimulus properties - Model comparison

1a. Eccentricity

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	2.747e -30	1.099e -29	1.000	
StDn-1	0.200	1.229e -29	4.918e -29	4.476	0.626
Eccentricity	0.200	0.004	0.015	1.385e +27	1.303
StDn-1 + Eccentricity	0.200	0.943	65.900	3.432e +29	3.951
StDn-1 + Eccentricity + StDn-1 ✻ Eccentricity	0.200	0.053	0.226	1.945e +28	2.392

1b. Spatial location (hemifield)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.232	1.208	1.000	
StDn-1	0.200	0.481	3.702	2.073	0.464
Hemifield	0.200	0.079	0.343	0.340	1.690
StDn-1 + Hemifield	0.200	0.181	0.883	0.780	3.545
StDn-1 + Hemifield + StDn-1 ✻ Hemifield	0.200	0.028	0.114	0.120	1.387

1c. Mean difference (n-1, n)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	3.111e -6	1.244e -5	1.000	
StDn-1	0.200	0.998	2550.135	320978.693	0.570
Mean difference	0.200	4.784e -9	1.914e -8	0.002	0.726
StDn-1 + Mean difference	0.200	0.002	0.006	502.187	0.815
StDn-1 + Mean difference + StDn-1 ✻ Mean difference	0.200	8.729e -7	3.491e -6	0.281	1.091

Note. All models include subject.

Table 1. Experiment 1. Serial dependence (associated with trial n-1) and stimulus properties. Each table presents the results of a Bayesian repeated-measures ANOVA on zREn, with two within-subject factors: StDn-1 and one property of interest: eccentricity, spatial location (peripheral blocks only: same or opposite hemifield than previous stimulus) and difference in the mean trajectories of the RDKs presented in consecutive trials. P(M): prior probability of each model, assumed to be equal for all. P(M/data): posterior probability of the model (given the data). BFM: Bayes factor for the model. BF10: Bayes factor for the alternative hypothesis relative to a null model (expressed by each model).

Positive serial dependence in variance extends up to the latest 2 trials

Investigations of serial dependence have typically focussed on the influence of very recent trial history, examining only the effect of the immediately previous and penultimate trials on reports. We examined serial dependence through trial history by

modelling the fixed-effects size of serial dependence while allowing for between-subject variability, building ten varying intercept, varying slope Bayesian linear mixed-effects models (LMM) with zRE_n as dependent variable, and StD_{n-t} (t=1…10, respectively) as independent predictor, with random-effects grouped by participant. We chose a uniform prior distribution over the real numbers for the fixed-effects coefficient and for the standard deviation of the by-subject varying intercepts and slopes, and a LKJ prior with shape parameter η= 2.0 for the random-effect correlation matrices. Unless stated otherwise, analogous priors were established for other Bayesian LMMs reported in this paper. Fixed-effects coefficient estimates were largely insensitive to prior selection, as can be seen in the example presented in the section 3.1 of the current chapter.

We applied these models to foveal and peripheral blocks separately as well as the overall dataset. Figure 2d presents the fixed-effect coefficients and 95% credible intervals for the association between past StD (up to trial n-10) and current report for all trials, as well as per eccentricity. The value of the LMM fixed-effects coefficient estimate for the effect of StD_{n-t} on zRE_n represents the linear slope for the relationship between the StD presented in trial n-t and the normalized response error provided in the current trial: in other words, the variation (in z-scores) in zRE_n when StD_{n-t} increases by 1º. Therefore, a positive B coefficient represents an attractive bias: a larger StD in a past trial drives a larger response in the present one, regardless of the current stimulus. Conversely, a negative B coefficient represents a repulsive bias.

The fixed-effects B coefficient estimates for the effect of StD_{n-1} and StD_{n-2} on zRE_n are positive, indicating an attractive bias. For StD_{n-1} (all trials pooled), B=0.0034 (0.0017, 0.0051) suggesting that, regardless of the value of StD_n, participants' judgments of visual variance increased by a magnitude of 0.0034 (z-score) per 1º increase in previous trial StD (StD_{n-1}). The effect of StD_{n-2} is weaker but still present: B=0.0014, (0.0003, 0.0026). To make clear the size of these effects, we can consider absolute responses as outcome variable (adding the current StD_n and the interaction with StD_{n-1} to the models). Here, the increase is of 0.0586 (0.0272, 0.0892) units per unit of StD_{n-1}, or an attractive effect

of 5.9% towards the previous stimulus, whereas for StD_{n-2} the effect size is 0.0242 (0.0006, 0.0483) (2.4%).

Thus, variance judgments at one specific trial (n) are attracted a small but meaningful amount towards the variance presented in the previous trial (n-1) and to a lesser extent the trial before (n-2). Note that, since the initial position of the response bar is randomized for each trial, simple motor routines involved in response execution cannot explain this serial dependence.

2.2.3. Variance reports are subject to a negative bias driven by less recent trial history

Looking past the previous 2 trials, as shown in figure 2d, a reversal from positive (trial n-1 and n-2) to negative B coefficient values is observed for less recent presentations, indicative of a negative (i.e. repulsive, anti-Bayesian) bias: current responses were *less* similar to the StD presented in those trials, in a manner akin to sensory adaptation after-effects (Kohn, 2007; Payzan-LeNestour et al., 2016). This effect started at trial n-4 and peaked at trials n-7 to n-9 (StD_{n-8}: B=-0.0021 (-0.0032, -0.0010)). Similar effect sizes and timescales are observed for foveal and peripheral presentations.

In Figure 2e (inset in 2d) we present a further exploration of the evolution of the negative effect in relation to more remote positions in trial history, up to trial n-30. Evidence for the negative effect shows a slow decline after trial n-9 but persists to some extent until trial n-20, disappearing for more remote trial positions.

To confirm that the observed serial effects were truly dependent on trial history, we conducted extensive control analyses exploring potential 'serial dependencies' in relation to random subsets of the data, future presentations (StD_{n+1}) and to shuffled data – see below, section 3.2 of the current chapter, 'Control Analyses'. These analyses

confirm that only in the true past trial history is there evidence for these serial dependencies, supporting that they are not simply due to statistical artifacts.

3. EXPERIMENT 1: SUPPLEMENTARY ANALYSES

3.1. BAYESIAN STATISTICS: FIXED-EFFECTS PRIOR AND PARAMETER ESTIMATION IN BAYESIAN LMM

In all the Bayesian varying-intercepts, varying-slopes linear mixed-effects models (LMM) reported in the previous section (and elsewhere in this thesis) we employ the same set of priors: a uniform prior distribution over the real numbers for the fixed-effects coefficients and for the standard deviation of the by-subject varying intercepts and slopes, and a LKJ prior with shape parameter $\eta = 2.0$ for the random-effect correlation matrices. However, first we tested several priors in order to assess the robustness of parameter estimation regarding prior selection. This section presents a single example, regarding a Bayesian by-subject varying-intercepts, varying slopes LMM for zRE_n as dependent variable, with StD_{n-1} as independent. Figure 3 presents the value of the fixed-effects B coefficient estimate for StD_{n-1} and its 95% credible intervals as a function of several manipulations on the width (main figure) or the mean (inset figure) of its prior. The priors for the standard-deviation of the by-subject varying intercepts and slopes and for the random-effect correlation matrices are kept constant in all cases, with the shapes specified above.

As for the fixed-effects prior, considering the effect of its width, $\sigma_{prior,}$ on parameter estimation for the fixed-effects B_{StDn-1} coefficient, we employed a flat prior over the real numbers and five different Gaussian priors with mean $\mu_{prior}=0$ and σ_{prior} that took the following values: 1, 0.1, 0.05, 0.01, 0.001. In other words, we tested six priors centred

around zero (indicative of absence of effect), each one more narrow than the previous, starting with a uniform prior, which could be approximated with a Gaussian prior with the general shape $N(0, \sigma_{prior})$, where σ_{prior} is infinite. The corresponding parameter estimates, depicted in the larger graph, show that parameter estimation is highly robust to manipulations of the width of the prior, unless the latter is forced into a very narrow shape, in which case the parameter estimate is strongly biased toward the peak of the distribution (toward zero).

Considering the effect of the mean of the prior (μ_{prior}) on parameter estimation, we tested three cases of the general shape $N(\mu_{prior},1)$, with μ_{prior} taking the values -1,0,1. Shifting the prior mean (which in the case of Gaussian shape coincides with its peak) toward negative and positive, respectively, implies biasing the evidence in favour of (*a priori*) negative and positive serial dependence, respectively. However, as shown in the inset of figure 3, these shifts (of 1 unit toward negative or positive) had very little effect on parameter estimation with a prior width equal to the amount of shift: $\sigma_{prior}=1$.

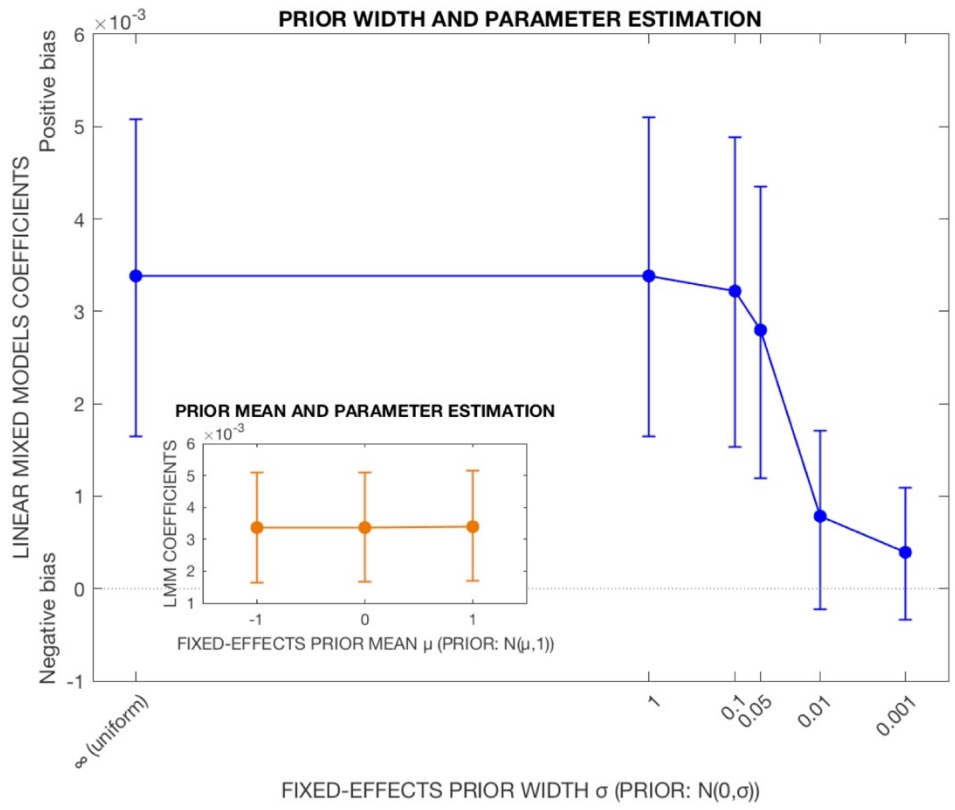


Figure 3. Bayesian statistics. Influence of prior selection on fixed-effects parameter estimation in a Bayesian LMM**.** The main figure shows the effect of manipulations of the width of the prior, which has a general shape of N(0, σ_{prior}), i.e. it's centred around zero in all cases, with σ_{prior} ranging from 'infinite' (actually a flat prior over the real numbers) to 0.001. The distances in the horizontal axis are on a logarithmic scale (the value corresponding to the flat prior is at the distance for $\sigma_{prior}=10^6$. The plot shows that parameter estimation is highly robust to changes in the width of the prior up to $\sigma_{prior}=0.1$, becoming strongly attracted to the prior mean when narrower priors are established. The figure inset shows the effect of manipulations on the mean of the prior on parameter estimation. In all three cases, priors take the general shape of N(μ_{prior}, 1), with μ_{prior} taking the values -1/0/1 (i.e. a shift of 1 unit toward both directions around zero, equal in magnitude to the width of the prior,1). Results show that parameter estimation is virtually insensitive to such shifts in the prior mean.

3.2. EXPERIMENT 1: CONTROL ANALYSES

In order to demonstrate that the observed serial effects are truly dependent on trial history and not due to statistical artifacts in the experimental design, we performed

three control analyses on Experiment 1 data: future-related serial dependence, serial dependence with shuffled responses and serial dependence within random subsets of the data.

3.2.1. Serial dependence in relation to future (n+1) trials

Regarding the first approach, we assessed whether there was any association between the StD presented in trial n+1 and the current response zRE_n. Naturally, future presentations should not be able to affect current perceptual judgments, so any supposed 'effect' found for StD_{n+1} should signal the presence of statistical artifacts in our data. We analysed StD_{n+1}-related 'serial dependence' by two methods: (i) by running a Bayesian repeated-measures (RM) ANOVA on the influence of StD_{n+1} level (as within-subject factor) on zRE_n (equivalent to the analyses conducted in the main text for StD_{n-1}) and (ii) by running a Bayesian linear mixed-effects model (LMM) for zRE_n (dependent variable), with StD_{n+1} as independent variable (random-effects grouped by participant's ID). Additionally, we repeated both analyses in relation to StD_{n+2}.

Figure 4a and 4b present the average zRE_n as a function of StD_{n+1} and StD_{n+2}, respectively, for all trials pooled, as well as for foveal and peripheral trials separately. The flat plots strongly suggest that there is no effect of StD_{n+1}/ StD_{n+2} on zRE_n, as expected. In a Bayesian RM ANOVA for zRE_n with StD_{n+1} as within-subject factor, the best model according to the analysis was the null (containing ID only: it is the reference model with $BF_{10\ null}$=1.000). The Bayes factor for the model containing StD_{n+1} (and ID) was $BF_{10\ StDn+1}$=0.042; therefore, there was strong evidence against any explanatory effect of StD_{n+1} compared to the null model ($BF_{null/StDn+1}$=1/0.042=23.810). Likewise, for the RM ANOVA with StD_{n+2} as within-subject factor, evidence in favour of the null was very strong: $BF_{null/StDn+2}$=1/0.031=32.258. In other words, neither StD_{n+1} nor StD_{n+2} had any effect on zRE_n. Figure 4c plots zRE_n as a function of StD_{n-8}, StD_{n-1} and StD_{n+1}, showing the difference between the slopes of each plot: the descending and ascending plots for

StD_{n-8} and StD_{n-1} indicate a negative (repulsive) and a positive (attractive) effect, respectively, whereas the flat line for StD_{n+1} indicates no effect.

Figure 4d presents the B coefficients and 95% credible intervals for the effect of StD_{n+1} and StD_{n+2} on zRE_n, according to two Bayesian LMMs. They have virtually zero values: $B=2*10^{-5}$ (-0.0009, 0.0010) for StD_{n+1} and B=-0.0001 (-0.0011, 0.0009) for StD_{n+2}, again confirming the absence of 'serial dependence' by trials n+1 and n+2.

3.2.2. Serial dependence in relation to shuffled datasets

As another 'sanity check' for confirming that the observed effects are truly dependent on trial history, we generated shuffled datasets from each participant's data, by permuting the responses provided by each participant within each StD level. In other words, each simulated dataset had exactly the same trial sequence as the real one, and the responses given for each StD level were identical as the participant's responses, but they were randomly assigned to any trial for the same participant and StD level. This means that everything was kept the same as in the experimental data, except for the relation between each response zRE_n and its trial history. Therefore, if we failed to find serial dependence in these simulated datasets, we would need to conclude that the observations in our data cannot arise from anything else than the specific trial history leading to each zRE_n.

We generated 10 shuffled datasets (each encompassing the shuffled data from all 30 participants in Experiment 1) and ran 10 Bayesian LMMs for each of them, assessing the influence of StD_{n-t} (t=1...10) on the shuffled zRE_n. The average and standard errors of the obtained coefficients are presented in figure 4d together with the coefficients found for the real data. It can be observed that there is no serial dependence for the shuffled datasets.

3.2.3. Serial dependence in random data subsets

Finally, we examined serial dependencies in the real, past trial history, but within a random subsample of trials instead of the entire dataset. By this control analysis we aimed to rule out any artifact due to the trial sequence being a closed set with a fixed number of pseudorandomized, equally frequent StD presentations. Nevertheless, such an artifact would have appeared equally for future and past trials, an issue that has already been discarded; this third control serves as a mere confirmation. Note that, while for future trials and shuffled datasets we would expect true serial dependencies to *disappear*, in the current control the 'sanity check' requires the positive and negative effects to *persist* in the subset of trials.

We obtained twenty subsets of the Experiment 1 dataset by randomly selecting half of the trials of each experimental block (30/60). Note that this subsampling pertains to the 'current trials', for which the effect of their true trial history was analysed; in other words, the trial history of the selected items was not altered by the subsampling.

Figure 4e summarizes the result of analysing serial dependences, up to trial n-10, in these twenty subsets. Each data point of the plot presents the average of the LMM B coefficient estimates of 20 Bayesian linear-mixed effects models for the influence of StD_{n-t} (t=1 … 10) on the current normalized response error, zRE_n, with each of the 20 models ran on a different subset. The error bars represent the standard error. The ten data points correspond to the ten points of trial history (n-t, with t=1 … 10) at which serial dependencies are analysed for all 20 subsets. Figure 4e show that both the positive and negative after-effects persist in similar magnitude and timescale than for the entire dataset.

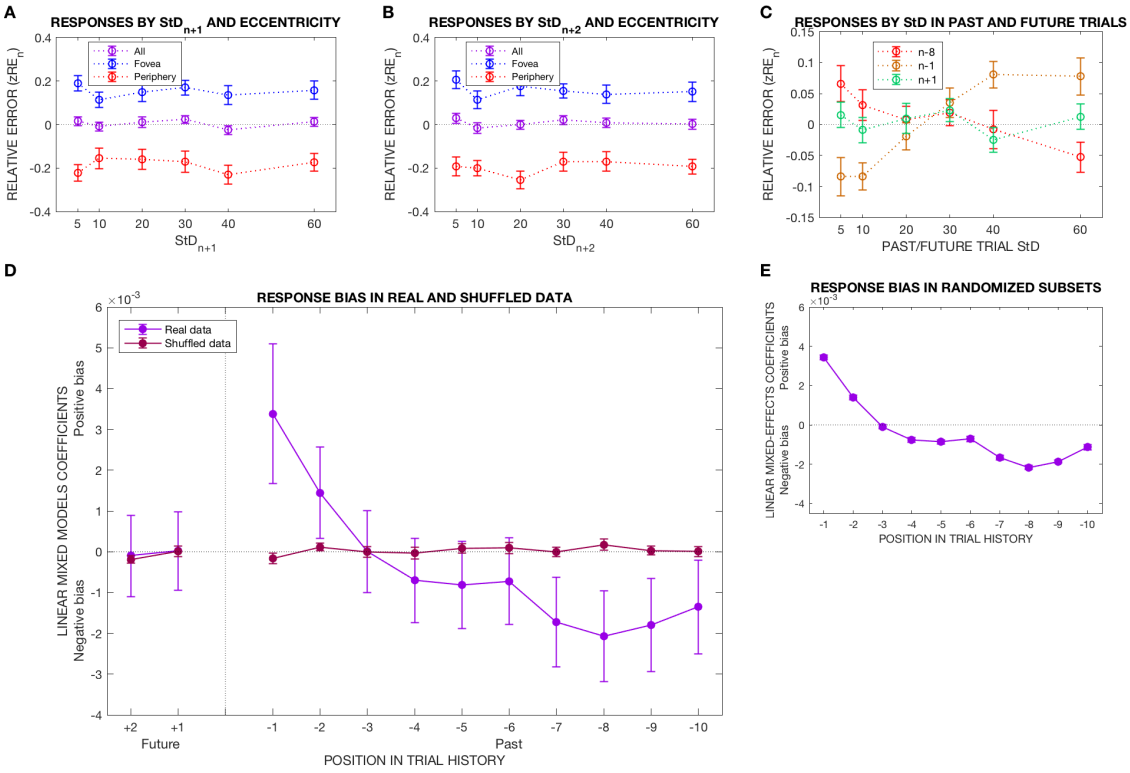


Figure 4. Experiment 1: Control analyses. 4a-4b. Normalized relative error in current response (zREn) as a function of the StD presented in the following trial (StDn+1) (**4a**) and in trial n+2 (**4b**), plotted by eccentricity: all trials pooled, foveal and peripheral trials separately. In all cases, the approximately flat plots indicate that future trials StD have no effect on current response (as expected). The error bars represent the between-subject standard error. **4c.** Normalized relative error in current response (zREn) as a function of the StD presented in trial n-8, n-1 and n+1. The slopes of past trial plots indicate a negative and positive bias in relation to n-8 and n-1, respectively, whereas the flat slope for n+1 once more indicates lack of effect. The error bars represent the between-subject standard error. **4d.** Serial dependence in real and shuffled data. The purple plot presents the fixed-effects coefficient estimates and 95% credible intervals in 12 Bayesian linear mixed-effects models (LMM) with zRE_n as dependent variable and the StD presented in trials n+1, n+2, and n-1 … n-10 as independent variable, respectively for each model. The red plot represents shuffled data: simulated datasets where everything has been kept the same as in the real experiment except for the specific association between each zRE_n and its corresponding trial history, which has been shuffled. We generated 20 shuffled datasets and ran the 12 LMMs on each of them. Thus, each datapoint represents the average of the fixed-effects coefficients obtained for the 20 datasets at each trial position, and the error bars indicate the standard error for the results obtained in the 20 datasets. The plots show that, unlike in the real data, removing the true association between each response and its trial history eliminates the observed serial dependence in relation to past trials, whereas there is no serial dependence in relation to future trials, neither in the real nor in the shuffled data. **4e.** Serial dependence in twenty random subsamples of the real data. The plot summarizes the result of 20 iterations consisting on randomly subsampling half of the trials of each participant's data and running 10 Bayesian LMMs for the influence of StD_{n-t} (t=1 … 10) on zRE_n, applied to the selected subset. Each datapoint represents the

average and standard error of 20 LMM B coefficient estimates for the effect of StD_{n-t} on zRE_n, observed at different points in trial history (n-t, t=1 … 10), separately for the 20 random subsets. The pattern of serial dependencies is almost identical to the one observed for the entire dataset, further confirming that serial dependencies do not arise only as a result of statistics being performed on a closed set of StD presentations.

3.3. EXPERIMENT 1B: ENSEMBLE MEAN AND SERIAL DEPENDENCE IN VARIANCE

In Experiment 1, the similarity of the mean trajectories of consecutive RDKs did not affect the size or sign of the serial dependence in variance reports (see section 2.2.2 of the current chapter). This result is in line with other findings in our study that suggest a high-level source of such serial bias (such as independence of eccentricity and spatial location). However, it may seem counter-intuitive given the close relationship between both statistics in ensemble vision, as well as the purported function of serial dependence, namely promoting perceptual continuity: it seems reasonable to expect that the bias on a certain dimension should be stronger if consecutive stimuli share similar attributes.

As in Experiment 1 mean trajectories could take any integer value from 0^o to 359^o, we decided to run an additional experiment (1B) with a limited number of possible means, to rule out that our failure to find any effect of mean similarity on serial dependence in variance was related to the employment of such wide range of trajectories. It could be the case that the attractive bias was much stronger when mean directions were almost identical but became mean-invariant over a certain amount of divergence. Alternatively, the great number of possible mean trajectories could render it futile to rely on trial-by-trial mean estimation to compute variances.

In addition to serial dependence, we wanted to assess the interaction between mean similarity and other aspects of the task, such as accuracy and precision in variance

estimations.

3.3.1. Methods

The methodology was similar to Experiment 1, although in Experiment 1B stimulus presentation was always foveal. Regarding the mean direction of the RDKs, instead of being randomized among all integers from 0º to 359º, only three values were allowed, presented with equal frequency: 10º, 45º and 100º (0º corresponds to the rightward horizontal axis and the increase proceeds clockwise). Thus, the difference between the mean trajectories of two trials could take four values: 0º, 35º, 55º and 90º. Each experimental session was formed of 240 trials.

3.3.2. Results

Twenty-one participants (11 female, mean age 20.5 y/o, standard deviation 2.13) took part in the study, three being members of the laboratory and the rest paid volunteers recruited through the SONA system and by online advertisement in the University website. The overall number of trials was 5040.

Mean similarity does not affect performance on variance judgments

The central variable in this experiment was the absolute inter-trial mean difference, specifically the difference between RDK mean in the current and previous (n-1) trials: $\Delta\mu_{n,n-1} = |\mu_{n-1} - \mu_n|$, as we considered that, if the similarity of RDK means had any effect on sequential variance judgments, this effect would be particularly meaningful regarding consecutive RDK presentations.

To ascertain whether $\Delta\mu_{n,n-1}$ affected performance on iterative variance estimation (in other words, whether mean similarity exerted priming effects on variance judgments), we analysed its influence in response time (RT), accuracy and precision. For both accuracy and precision we employed inverse measures: inaccuracy or error size and imprecision or response dispersion. Response (in)accuracy, or 'error size', was defined as the absolute value of the difference between response and veridical StD: $E_n = | StD_n - R_n |$. Response (im)precision or dispersion was computed as the standard deviation of the responses provided by each participant for a certain combination of StD_n and inter-trial mean difference (σ_R). For all three measures (RT, error size and response dispersion), smaller values indicate better performance. A Bayesian repeated-measures ANOVA was conducted on the effect of current StD (StD_n) and inter-trial mean difference ($\Delta\mu_{n,n-1}$) –as within-subject factors- on each of these three measures of performance (RT, error size and dispersion). Thus, we conducted three RM ANOVAs with the same two within-subject factors, but with a different dependent variable for each. In each case, a model comparison was performed based on the results of the Bayesian RM ANOVA, assessing the evidence in favour of each possible combination of independent variables in terms of explaining the variability of the performance measure. The five competing models, in each case, were the null model, a model with StD_n as the sole independent variable, a model with only $\Delta\mu_{n,n-1}$, a model with both main effects (StD_n and $\Delta\mu_{n,n-1}$) and the full model with both main effects and the interaction term $StD_n * \Delta\mu_{n,n-1}$. If intertrial mean difference had any role in performance in the current trial (measured by RT, error size or response dispersion), a model containing $\Delta\mu_{n,n-1}$ would be more explanatory than a model without this variable.

In all three ANOVAs, the best model was the one containing StD_n only, with a Bayes factor of $BF_{10}=2.352*10^6$ (RT); $4.658*10^{67}$ (error size); $2.276*10^{20}$ (dispersion). This model outperformed the second best (the model with both main effects, StD_n and $\Delta\mu_{n,n-1}$, in all three cases) by a factor of $BF_{StDn/main\ effects}$=55.79 (RT); 3.402 (error size); 34.344 (dispersion). The Bayes factor for inclusion of the variable of interest, $\Delta\mu_{n,n-1}$, was $BF_{inclusion}$=0.012 (RT); 0.294 (error size); 0.019 (dispersion), indicating moderate evidence

against any explanatory role of mean difference with regards of error size, and strong evidence against it with regards of RT and response dispersion. In summary, evidence indicated that inter-trial difference in RDK mean does not have any effect on performance in variance judgments, measured in terms of response time, accuracy or precision.

Mean similarity does not affect serial dependence by previous StD on variance judgments

In order to analyse the influence of mean similarity between consecutive trials on serial dependence in variance, first we conducted a Bayesian RM ANOVA on the effect of previous trial StD (StD_{n-1}) and inter-trial mean difference ($\Delta\mu_{n,n-1}$) –as within-subject factors- on current normalized response (zRE_n: normalized response error in variance judgments, as dependent variable). A model comparison was performed between all combinations of tested factors, based on the evidence given by the RM ANOVA: the five competing models were the null model, two models with a single main effect each (StD_{n-1} only, $\Delta\mu_{n,n-1}$ only), a model with both main effects (StD_{n-1} and $\Delta\mu_{n,n-1}$) and the full model with both main effects and the interaction term (StD_{n-1}, $\Delta\mu_{n,n-1}$ and $StD_{n-1} * \Delta\mu_{n,n-1}$). If mean similarity had any influence of serial dependence, it must be able to modulate the effect of StD_{n-1} on zRE_n: consequently, the model including the interaction term $StD_{n-1} * \Delta\mu_{n,n-1}$ should be more explanatory than any other model.

According to the results of the Bayesian RM ANOVA, the best model contained both main effects, StD_{n-1} and $\Delta\mu_{n,n-1}$, but not the interaction; this model outperformed the full model (including the interaction term, key to our hypothesis) by a factor of $BF_{main\ effects/full}$=69.325, indicating strong evidence in favour of the main effects compared to the full model. Overall, the evidence for inclusion of the interaction term was $BF_{inclusion}$=0.050, indicating strong evidence against this term. In conclusion, while inter-trial mean difference apparently drives a systematic response bias on variance

judgments (as shown by the fact that the best model includes the main effect of $\Delta\mu_{n,n-1}$), there is strong evidence in support of the absence of effect of inter-trial mean difference on StD_{n-1}-related serial dependence in variance.

CHAPTER 2: EXPERIMENT 2: PROCESSING STAGES INVOLVED IN SERIAL DEPENDENCE IN VARIANCE REPORTS

This chapter presents the results for a series of experiments (collectively referred to as Experiment 2) wherein different manipulations to the stimulus and task of Experiment 1 were applied in order to investigate which levels of perceptual processing give rise to the serial dependencies described in the previous chapter, in particular to the positive effect driven by very recent history. According to our results, positive serial dependence by a past recent trial was unaffected by whether or not a response had been performed in that trial (Experiment 2A), but disappeared entirely if participants had been required to produce a decision about a different feature-dimension of the stimulus (mean instead of variance), regardless of whether the required decision had been pre-cued (Experiments 2B-2C) or post-cued (Experiment 2D). This showed that serial dependence in variance arose as a result of past, dimension-specific perceptual decisions, and not merely attentive perception. Results of these experiments as well as Experiment 2E (which applied modifications to different time intervals within the trial structure) suggested that serial dependence was subject to time and (dimension-specific) capacity manipulations, similar to those observed in working memory operations. Experiments 2A and 2B have been published in the Journal of Vision as part of the article 'Serial Dependence in the Perception of Visual Variance' (Suárez-Pinilla et al., 2018b).

In the previous chapter we reported evidence for serial dependence in judgments about the variance of RDK stimuli. Specifically, we found two opposite types of bias at different timescales: an attractive, Bayesian-like bias related to the StD of the very recent (n-1 and n-2) trials and a repulsive, negative bias which operates on a longer timescale.

At what level of processing do these serial dependencies exert their influence? The non-local nature and independence from inter-trial similarity in RDK direction mean suggest that attractive serial dependence may not be driven by low-level, sensory processes. However, the specific stages of variance processing at which it arises are yet to be determined (Cicchini et al., 2017; Fischer & Whitney, 2014; Fritsche et al., 2017; John-Saaltink et al., 2016; A. Kiyonaga et al., 2017). To address this issue, we designed a new series of experiments (collectively referred to as Experiment 2) where we applied several manipulations to the task to disambiguate the contributions of low-level sensory processes, perceptual decisions and responses to serial dependence of variance judgements.

Experiment 2A aimed to isolate the contribution of response to the serial dependence effect by introducing 'no-response' trials, to exclude the influence of physically making a response. However, in order to prevent attention detachment in no-response trials, such trials were not pre-cued, meaning that potential contribution of decision-making and response preparation during stimulus presentation could not be ruled out.

For this reason, in Experiment 2B we employed a (pre-cued) task-switching design to disentangle the contribution of perception and decision processes. In this experiment, participants had to report either the mean or the variance of the RDK motion direction, in order to disambiguate the effect of past perceptual history (as all presented RDKs have a certain variance) and past decisions about variance (made only in certain trials). The required decision in each trial (mean or variance) was pre-cued to discourage participants from starting to decide about both feature-dimensions before the response phase of the trial, which might have rendered the experiment futile. However, pre-cuing

likely caused a difference in feature-based attention during stimulus presentation. Since serial dependence is enhanced by perceptual attention (Fischer & Whitney, 2014), a potential difference between conditions, suggestive of an effect of decision processes on serial dependence, may rather be explained by low-level, attention-modulated perceptual processes.

To rule out a confounding effect of feature-based attention, we performed another task-switching experiment wherein the required decision was post-cued (after stimulus offset). In this scenario, any difference in serial dependence in function of the decision required in trial history must be of post-perceptual origin. In order to ameliorate the problem of participants preparing both decisions at the same time during stimulus presentation, we shortened the duration of the RDK to 250 ms (it was 500 ms in previous experiments). Consequently, for a strict comparison of the pre-cued and post-cued paradigms, we repeated the pre-cued design with shorter RDKs (250 ms); this design is referred to as Experiment 2C. The post-cued task-switching (with 250-ms RDK) is reported as Experiment 2D.

As detailed in the corresponding sections below, results from Experiments 2A-2D suggest that (positive) serial dependence in variance judgments is driven by dimension-specific decision-making, rather than by response, low-level sensory processes or perceptual attention. We therefore enquired how decisional representations may be stored and implemented to affect future judgments. If the process involved some kind of 'memory' (such as, for example, working memory representations implemented for decision-making (Bliss et al., 2017; A. Kiyonaga et al., 2017)), we may observe an effect of time and capacity limitations on serial dependence. We therefore devised a final experiment (Experiment 2E), where we combined the task-switching design of Experiments 2B-2D with manipulation of different time intervals at several points of the trial structure.

1. EXPERIMENTS 2A-2B: EFFECTS OF RESPONSE AND DECISION ON SERIAL DEPENDENCE

1.1. METHODS

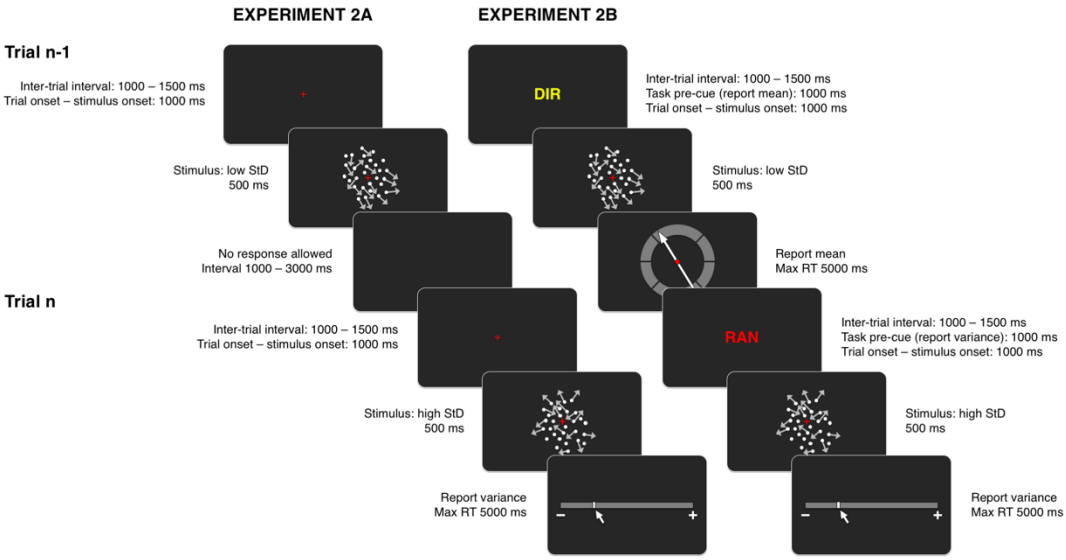


Figure 5. Experiments 2A-2B: Methods. Each trial presented an RDK of a certain mean and variance (standard deviation, StD) in the motion trajectories of its component dots. In the example, trial n-1 and n have low and high StD values, respectively. The stimulus was presented for 500 ms and always in the fovea (0 dva). Both experiments interleaved 2/3 of trials in which variance reports were required, and 1/3 in which either no response (2A) or mean trajectory estimation (2B) was required. In 2B the trial type was pre-cued, so that the word DIR and RAN at the beginning of each trial indicated whether a mean or variance judgment was required for that trial. Abbreviations: StD – standard deviation (of the RDK motion directions), RT: response time, ms: milliseconds.

Figure 5 presents an overview of the experimental structure of Experiments 2A and 2B. The methods of these experiments were similar to those of Experiment 1 with the following exceptions:

1.1.1. Stimuli

All stimuli were presented on the centre of the screen (where a fixation cross was displayed) and eye-tracking was not performed, as visual eccentricity was not under examination.

1.1.2. Procedure

Experiment 2A had a practice block with 72 trials and ten experimental blocks with 60 trials each; same for 2B except that the practice block was longer (90 trials), due to the additional demands of its task-switching design (see below). In Experiment 2B 6 participants (out of 15) performed a session half as long (5 blocks), due to different availability of diverse participants. For both experiments, in each block, 2/3 of the trials required 'randomness' scores as described for Experiment 1. In Experiment 2A, the remaining 1/3 were no-response trials: after stimulus presentation, instead of the response bar only a blank screen appeared for a randomized interval between 1000-3000 ms, after which the next trial started. Participants were told in advance that they should expect a certain number of no-response trials, but they did not know the proportion and these trials were not pre-cued in any way. In Experiment 2B, 1/3 of trials required participants to report the 'mean' direction of the motion of the RDK, by adjusting a rotating arrow with the mouse (see Figure 5). The required task was pre-cued at the beginning of the trial: one three-letter word, either 'RAN' ('randomness' report required) or 'DIR' (mean direction report required) was displayed for one second before the appearance of the fixation cross. The rest of the trial structure was the same as in Experiment 1 (only the response scale differed in RAN and DIR-trials).

1.2. RESULTS

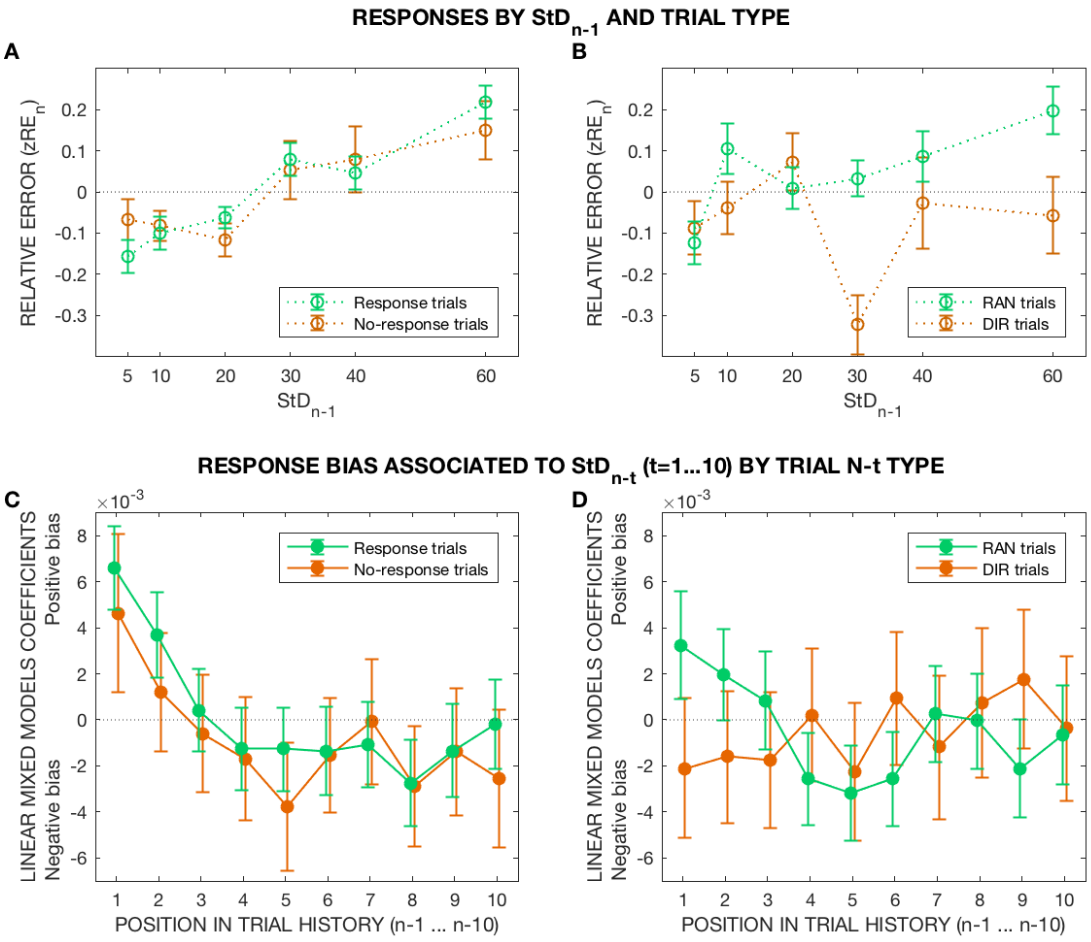


Figure 6. Experiments 2A-2B: Results. 6a - 6b. Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by trial n-1 type: response versus no-response in **6a** (experiment 2A), RAN versus DIR in **6b** (experiment 2B). The error bars represent the between-subject standard error. Both response and no-response trials are associated with a positive bias by trial n-1 (as suggested by the ascending plot lines in 6a), whereas in figure 6b, only RAN trials elicit such positive serial dependence. **6c – 6d.** Fixed-effects coefficient estimates in 20 Bayesian LMMs with StD_{n-t} (t=1…10) as predictor of current response (zRE_n), modelled separately by trial n-t type: in **6c**, response versus no-response trials (experiment 2A); in **6d**, RAN versus DIR trials (experiment 2B). Since the dependent variable is the current variance ('randomness') judgment, trial n is always a response (6c) or a RAN (6d) trial. The error bars represent the 95% credible intervals for the true value of the coefficient.

1.2.1. Experiment 2A: effect of response execution on serial dependence in variance reports

Fifteen Psychology students (13 female, mean age 20.4 y/o, standard deviation 5.3) volunteered in exchange for course credits, under the conditions described previously. The total number of trials collected across all participants was 9000, out of which 3000 were no-response trials.

Serial dependence of previous StD is not affected by response processes

Figure 6a shows the distribution of normalized variance reports (zRE_n) as a function of the previous trial StD (StD_{n-1}) and type, i.e. whether n-1 had been a response or a no-response trial. The ascending and roughly parallel plots for each trial (n-1) type suggest that serial dependence in relation to StD_{n-1} was similar in magnitude and sign (i.e. attractive effect) regardless of whether trial n-1 was a response or a no-response trial. To formally test this observation, we conducted a Bayesian repeated-measures ANOVA on the effect of StD_{n-1} and trial n-1 type (as within-subject factors) on zRE_n. A comparison of all possible models based on the results of this analysis is shown in Table 2a. The best model includes only StD_{n-1} ($BF_{10}=2.386*10^6$). There was strong evidence against the inclusion of the interaction term StD_{n-1}*trial n-1 type: $BF_{inclusion}=0.051$. In a direct comparison between the main-effects model and the full model the ratio was given by $BF_{main\ effects/full}=10.75$. This lack of interaction confirmed that StD_{n-1} effect on current report was independent of response execution.

Figure 6c shows the fixed-effects coefficient estimates and 95% credible intervals for 20 Bayesian LMMs for zRE_n, with StD_{n-t} (t=1…10) as putative predictor, split by trial n-t type and modelled separately. A similar pattern in terms of effect size and direction can be seen regardless of whether previous trials required response or not: an attractive bias

in relation to the latest two trials (weaker for n-2), a roughly zero effect of trial n-3 and a reversal toward a negative effect peaking around trials n-5 to n-9, with a similar magnitude and timescale than for Experiment 1. In conclusion, our results indicate that serial dependence does not arise from past responses.

1.2.2. Experiment 2B: effect of decision on serial dependence in variance reports

Experiment 2A demonstrated that serial dependence in visual variance is not due to response execution; however, as trials were not pre-cued as to whether a response would be required, these results do not disambiguate between perception and decision-making (response preparation). Therefore, in Experiment 2B we deployed a pre-cued task-switching design in which participants needed to prepare and respond to two different perceptual tasks in different trials: reporting the variance (RAN trials) or the mean (DIR trials) of the motion of the RDK.

Fifteen first-year Psychology students (13 female, mean age 21.4 y/o, standard deviation 8.8) participated in this experiment in exchange for course credits, under the conditions described above. In total they performed 7200 trials, out of which 2400 were DIR-trials (alternative task).

Serial dependence is related to dimension-specific decision-making

We analysed the data in similar manner to Experiment 2A, ascertaining the influence of trial type in the observed serial dependence on variance judgments. Figure 6b presents the distribution of variance reports (zRE_n) as a function of StD_{n-1} and trial n-1 type -i.e. whether it required a decision about variance (RAN) or mean (DIR). Only when successive decisions were both regarding variance do we see an ascending slope in

relation to increasing StD_{n-1}, suggesting that the attractive bias associated with StD_{n-1} was only exerted if a decision on that dimension had been made.

Table 2b presents a Bayesian repeated-measures ANOVA with StD_{n-1} and trial n-1 type (RAN/DIR) as within-subject factors. The most explanatory was the full model including both main effects and their interaction (BF_{10}=48.459), although the evidence in its favour compared to the model with only the main effects was anecdotal ($BF_{full/main\ effects}$ =2.026). However, evidence in favour of the interaction term was larger when taking in consideration all possible models: $BF_{inclusion}$=5.371 –moderate evidence for inclusion. Thus, results point to serial dependence by StD_{n-1} being dependent on which dimension participants had to judge in the previous trial.

As in previous experiments, we also examined serial dependence within a broader span of trial history. Figure 6d presents the fixed-effects coefficient estimates and 95% credible intervals for the association between StD_{n-t} (t=1…10) and zRE_n, after splitting the dataset according to the trial type at each position: thus, the influence of RAN and DIR trials is modelled separately by 20 Bayesian LMMs. As expected from the previous analysis, the positive effect related to StD_{n-1} is only present when those trials required participants to report variance; this is also the case for $StD_{n-2.}$ As for the negative effect appearing at longer timescales, it is clearly present in RAN trials, while for DIR trials, although the credible intervals for the coefficient contain zero at all trial positions (likely due to the smaller number of DIR trials), the negative effect seems to appear as early as trial n-1 (B=-0.0021 (-0.0051, 0.0009)), peak at trial n-5 (B=-0.0023 (-0.0052, 0.0007) and decrease afterward. The appearance of a negative serial dependence regardless of the task suggests that it may be sensory in origin - an adaptation after-effect.

TABLE 2. Serial dependence and task requirements - Model comparison

2a. Response (Experiment 2A)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	3.575e -7	1.430e -6	1.000	
StDn-1	0.200	0.853	23.192	2.386e +6	0.393
Trial n-1 type (response vs. no-response)	0.200	5.551e -8	2.220e -7	0.155	0.772
StDn-1 + Trial n-1 type	0.200	0.135	0.622	376281.756	1.178
StDn-1 + Trial n-1 type + StDn-1 ✻ Trial n-1 type	0.200	0.013	0.051	35151.882	2.835

2b. Dimension-specific judgment (Experiment 2B)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.012	0.048	1.000	
StDn-1	0.200	0.023	0.095	1.964	0.360
Trial n-1 type (variance vs. mean estimation)	0.200	0.109	0.489	9.202	1.348
StDn-1 + Trial n-1 type	0.200	0.283	1.578	23.921	2.994
StDn-1 + Trial n-1 type + StDn-1 ✻ Trial n-1 type	0.200	0.573	5.371	48.459	2.705

Note. All models include subject.

Table 2. Experiment 2: 2A-2B. Serial dependence (associated with trial n-1) and trial n-1 type (task requirements in that trial). Each table section presents the model performance on Experiment 2A and 2B datasets, respectively, according to the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and trial n-1 type: response/no-response in Experiment 2A, RAN/DIR in Experiment 2B. P(M): prior probability of each model, assumed to be equal for all. P(M/data): posterior probability of the model (given the data). BF_M: Bayes factor for the model. BF_{10}: Bayes factor for the alternative hypothesis relative to a null model (expressed by each model).

2. EXPERIMENTS 2C-2D: DISAMBIGUATION BETWEEN DECISION AND PERCEPTUAL ATTENTION: SERIAL DEPENDENCE IN PRE-CUED AND POST-CUED TASK-SWITCHING

In Experiment 2B, positive serial dependence in variance judgments only arose when the previous trial required a decision about variance. This result suggests that serial dependence in variance depends on dimension-specific decision-making, but a low-level origin modulated by attention cannot be entirely ruled out since the task was pre-cued before stimulus presentation, likely affecting deployment of perceptual attention to the cued feature-dimension.

For this reason, we conducted two more experiments to explicitly compare pre-cued (2C) and post-cued (2D) decisions. Naturally, if the task was post-cued, attention to mean and variance during RDK presentation couldn't depend on the current trial type. Experiment 2C (pre-cued task-switching) was at its core the same as 2B, but with some methodological differences (detailed below) to facilitate a direct comparison with 2D.

2.1. METHODS

The methodology of experiments 2C and 2D (summarized in Figure 7) was similar to Experiment 2B but departed in several ways as specified below.

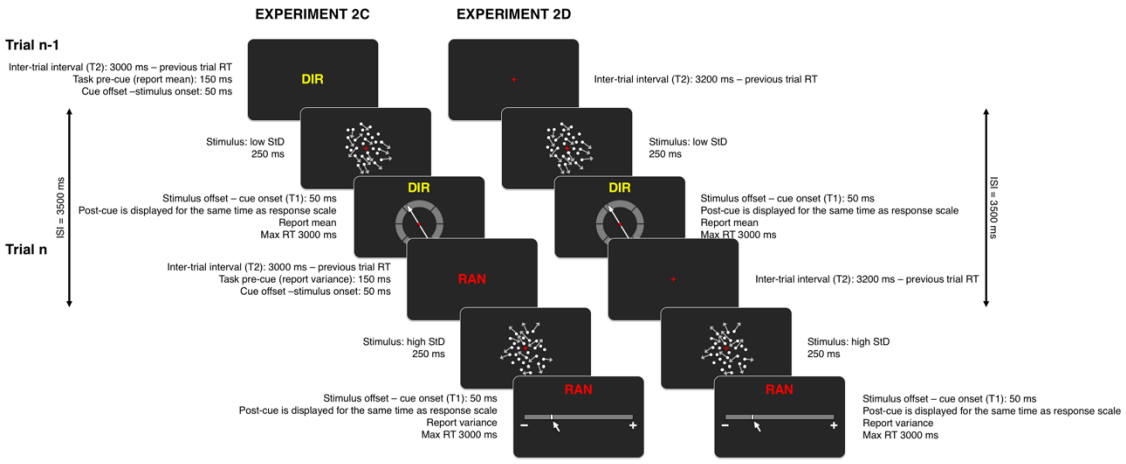


Figure 7. Experiments 2C-2D: Methods. Each trial presented an RDK of a certain mean and variance (standard deviation, StD) in the motion trajectories of its component dots. In the example, trial n-1 and n have low and high StD values, respectively. The stimulus was presented for 250 ms and always in the fovea (0 dva). In each trial, a three-letter cue (DIR or RAN) indicated whether a mean or a variance judgment was required; RAN and DIR trials were interleaved, with a frequency ratio 1:1. In Experiment 2C, the cue was displayed before (and after) stimulus presentation, while in experiment 2D it was only displayed after stimulus offset. The inter-trial time (interval between response offset and next trial onset, called T2) was adjusted in function of the response time, so that the inter-stimulus time (ISI: interval between consecutive RDKs onset) was kept constant at 3500 ms. Abbreviations: StD – standard deviation (of the RDK motion directions), RT: response time, ms: milliseconds.

2.1.1. Stimuli

For Experiments 2C and 2D the duration of the RDK was shortened to 250 ms (while Experiment 2B had used 500 ms). This was in order to prevent participants from starting a decision about both dimensions during stimulus presentation in the post-cued experiment.

2.1.2. Procedure

Both experiments 2C and 2D had a practice block of 80 trials and ten experimental blocks with 60 trials each.

In Experiment 2C, immediately at trial onset, a pre-cue was presented for 150 ms, indicating the task that the participant had to perform in that trial: either variance estimation (cued by word 'RAN') or mean estimation ('DIR'). After the cue disappeared, a blank screen with a central fixation cross was presented for 50 ms, and subsequently the RDK was displayed on the screen centre for 250 ms.

After another 50-ms interval (henceforth named T1), the response scale was presented along with the same cue as before: both items were displayed on the screen until the participant had made a response or the maximum time allowed (3000 ms) had passed. The response scale consisted on an analogous response scale for variance estimation or a rotating bar for mean estimation.

The inter-stimulus interval (time between onset of consecutive RDK stimuli, named ISI) was kept constant at 3500 ms regardless of the participant's response time, by adjusting the interval (henceforth T2) between response offset and next trial onset.

The trial structure for Experiment 2D was identical except that the stimulus was presented immediately at trial onset, without pre-cue. As in Experiment 2D, the post-cue was presented 50 ms after stimulus offset, together with the response scale. The maximum time allowed for response (3000 ms) and the constant inter-stimulus interval (3500 ms) were also the same as in experiment 2D. Due to the absence of pre-cue, even when the maximum time for responding was exhausted, there was at least a 200-ms interval before next trial onset.

Unlike Experiment 2B, where only 1/3 were DIR trials to maximize statistical power, in these experiments the frequency ratio of RAN and DIR trials was 1:1; both trial types were randomly interleaved within each block. By making both tasks equally frequent, we aimed to prevent any bias in dimension-specific attention due to different expectations for the frequency of each task in the post-cued experiment -although even if such bias had existed, it would affect equally both trial types.

2.2. RESULTS

2.2.1. Serial dependence is related to dimension-specific decision-making, also after excluding differences in dimension-specific perceptual attention by post-cued task-switching

Fourteen Psychology Students and members of the laboratory volunteered to each experiment, 2C and 2D (2C: all female, mean age 20.2 y/o; 2D: 9 female, mean age 21.2); one participant participated twice in both experiments. The dataset for each experiment had 9000 trials in total, of which half (4500) were of either type (RAN/DIR).

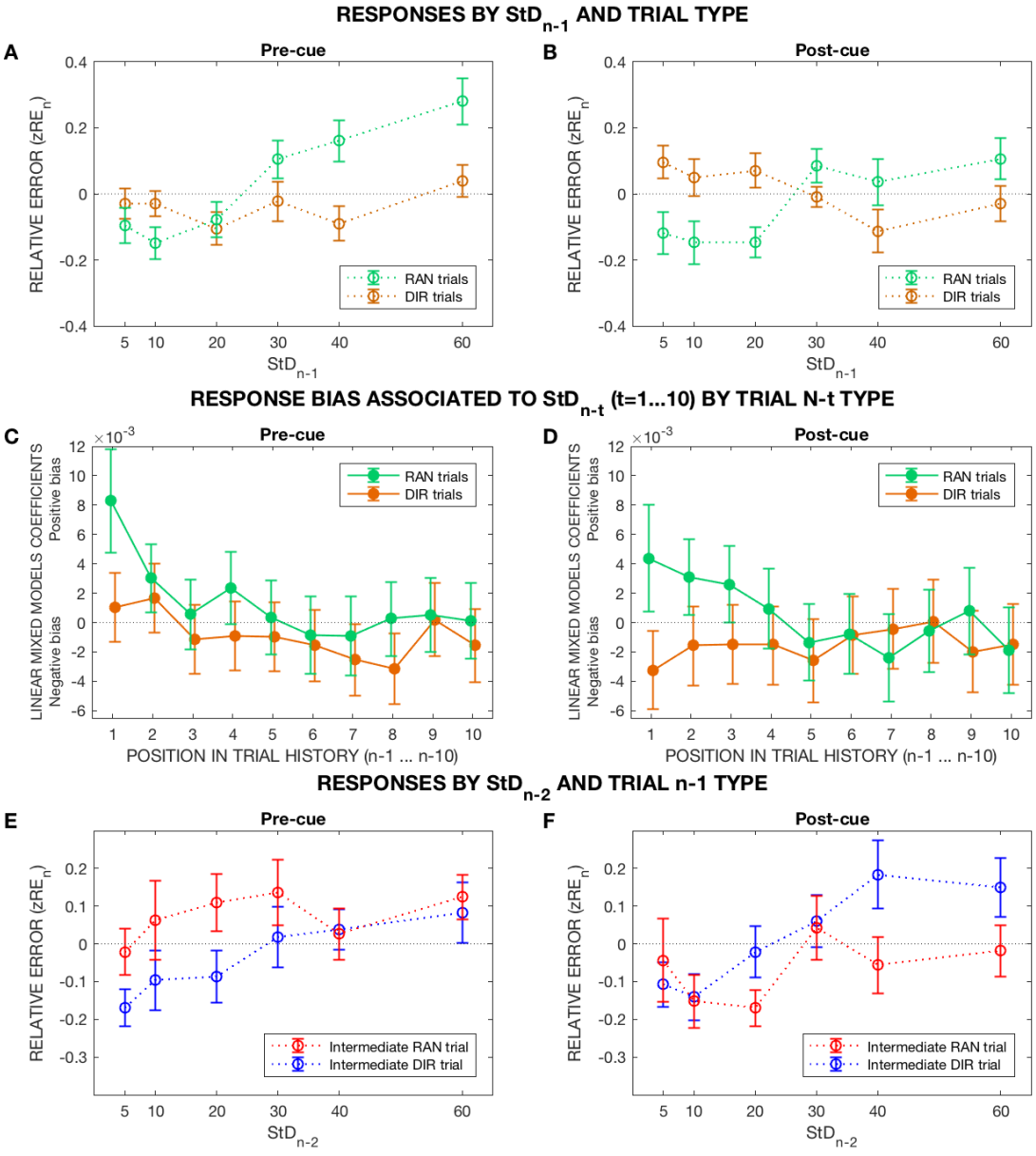


Figure 8. Experiments 2C-2D: Results. Comparison of serial dependence in a pre-cued versus post-cued task-switching paradigm: similar results in both experiments indicate that the differences between tasks are not attention-related. **8a - 8b.** Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by trial n-1 type: RAN (variance report required) and DIR (mean report required). In **8a**, the trial type (i.e. the required decision) is pre-cued, while in **8b** it is post-cued. The error bars represent the between-participant standard error. In both cases we observe positive serial dependence only for RAN n-1 trials, as suggested by the ascending slope of the RAN plot. No trace of positive serial dependence is observed in relation to post-cued n-1 DIR trials, indicating that the process that gives rise to serial dependence in variance is necessarily post-perceptual. **8c-8d.** Fixed-effects coefficient estimates in 20 Bayesian LMMs with StD_{n-t} (t=1…10) as predictor of current response (zRE_n), modeled separately by trial n-t type (RAN/DIR). Trial type is pre-cued in **8c** and post-cued in

8d. Since the dependent variable is the current variance ('randomness') judgment, trial n is always a RAN trial. The error bars represent the 95% credible intervals for the true value of the coefficient. Analyses further confirm that positive serial dependence only arises in relation to dimension-specific decision-making, even if such decision was post-cued. **8e-8f.** Normalized relative error in current response (zRE_n) as a function of the StD presented in trial n-2 (StD_{n-2}), plotted separately by trial n-1 type: RAN (variance report required) and DIR (mean report required). Both trial n-2 and trial n are RAN trials. The required decision is pre-cued in **8e** and post-cued in **8f**. In both cases, we observe a neat ascending slope for n-1 DIR trials. By contrast, when the intermediate trial (n-1) was a RAN trial, the effect of StD_{n-2} is less clear. In other words, positive serial dependence driven by a decision about variance made 2 trials before is eroded if another decision about the same feature-dimension (variance) is interposed (RAN n-1 trial), compared to cases where the intermediate decision was about a different dimension (DIR n-1 trial). This suggests an iterative modification of decisional representations as a new decision about the same feature-dimension is made, consistent with the hypothesis that sees serial dependence as result of the interaction of past and present decisional representations in visual working memory.

TABLE 3. Serial dependence and cued decision - Model comparison

3a. Pre-cued dimension-specific judgment (Experiment 2C)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	1.070e -4	4.281e -4	1.000	
StDn-1	0.200	0.029	0.121	273.553	0.364
Trial n-1 type (variance vs. mean estimation)	0.200	1.428e -4	5.711e -4	1.334	1.234
StDn-1 + Trial n-1 type	0.200	0.052	0.221	488.333	1.719
StDn-1 + Trial n-1 type + StDn-1 ✻ Trial n-1 type	0.200	0.918	44.913	8580.923	10.424

3b. Post-cued dimension-specific judgment (Experiment 2D)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.286	1.601	1.000	
StDn-1	0.200	0.014	0.058	0.050	0.490
Trial n-1 type (variance vs. mean estimation)	0.200	0.088	0.387	0.308	1.677
StDn-1 + Trial n-1 type	0.200	0.004	0.018	0.016	3.618
StDn-1 + Trial n-1 type + StDn-1 ✻ Trial n-1 type	0.200	0.607	6.186	2.125	1.526

Note. All models include subject.

Table 3. Experiment 2: 2C-2D. Serial dependence (associated with trial n-1) and cued decision in trial n-1. Each table section presents the model performance on Experiment 2C and 2D datasets, respectively, according to the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and trial n-1 type (pre-cued RAN/DIR trial in Experiment 2C, post-cued RAN/DIR trial in Experiment 2D). P(M): prior probability of each model, assumed to be equal for all. P(M/data): posterior probability of the model (given the data). BF_M: Bayes factor for the model. BF_{10}: Bayes factor for the alternative hypothesis relative to a null model (expressed by each model).

Figures 8a and 8b present the distribution of normalized variance reports (zRE_n) as a function of StD_{n-1} and trial n-1 type, for experiment 2C (pre-cue) and 2D (post-cue), respectively. On visual inspection we observe a similar pattern in both experiments, and similar as well to Experiment 2B (pre-cued task-switching with minor methodological differences compared to 2C). A roughly ascending plot is observed for current reports in relation to StD_{n-1} presentation, only when trial n-1 required a judgment about variance (RAN n-1 trial). On the contrary, we do not observe any trace of positive serial dependence in relation to StD_{n-1}, if the required decision in trial n-1 was about the mean of the RDK motion. Rather, there seems to be a negative effect associated to StD_{n-1} for post-cued n-1 DIR trials. Results in the post-cued design appear to confirm the post-perceptual origin of serial dependence, in relation to dimension-specific decision-making.

Table 3 presents the results of a Bayesian RM ANOVA on current variance report (zRE_n) with two within-participant factors: StD_{n-1} and trial n-1 type (RAN/DIR), computed on the data of Experiments 2C (pre-cued task-switching) and 2D (post-cued task-switching, respectively). In both cases, the model with more explanatory power is the one with all terms: both main factors and the interaction term StD_{n-1}*trial n-1 type, although for Experiment 2D the advantage over the null model is anecdotal (BF_{10}=2.125). Nevertheless, in Experiment 2D there is moderate evidence for inclusion of the interaction term StD_{n-1}*trial n-1 type, the one that determines whether the influence of StD_{n-1} (serial dependence) is different for each trial type: $BF_{Inclusion}$=6.135. Furthermore,

if the comparison is made between two identical models except for the inclusion of the term of interest (interaction term), namely between the model with both main factors and the one that also contains the interaction term, evidence in favour of the latter is extreme: $BF_{full/main\ effects}$=139.563.

Figures 8c and 8d depict serial dependence in relation to trials up to n-10: each data point represents the fixed-effects B coefficient for the relationship between StD_{n-t} (t=1…10) and current normalized variance judgment (zRE_n) according to a Bayesian LMM. Data are split by trial n-t type (RAN/DIR), while the current trial (n) is necessarily a RAN trial. Figure 8c corresponds to the pre-cued (2C) and 8d to the post-cued (2D) experiment. Once more we observe that the positive bias exerted by recent trial history is only driven by trials wherein a decision about variance was required, even if the cue for the required decision was presented after stimulus offset. Interestingly, this positive bias seems to last longer than in previous experiments: it appears still present in relation to trials n-3 or n-4, while in previous cases it had entirely disappeared by n-3. We propose two non-exclusive explanations for this: the interposing DIR trials (which are half of the total trials in the current experiments) might reduce the number of interposing decisions about variance, thus enhancing the carry-over effect of more remote decisions that are not disrupted by subsequent operations about the same feature-dimension (see below, section 2.2.2). In addition, the shorter duration of the RDK presentations (250 ms instead of the previous 500 ms) might render the competing negative effect weaker (particularly if it is related to adaptation processes).

In summary, our results show that positive serial dependence depends on dimension-specific decisions made in recent trials, ruling out alternative explanations such as the influence of response (2A) or differences in perceptual attention (2D). For equal stimuli, task requirements and attentional deployment during perception, as guaranteed by the post-cued design, we fail to encounter *any* trace of attractive serial dependence in relation to the variance of past trials that did not require a (post-perceptual) decision about variance.

2.2.2. Serial dependence by distant trials is disrupted by subsequent dimension-specific decision-making

Since we have established that serial dependence is an attractive effect exerted by past decisions, and the examination of distant trials shows a progressive decline and eventual disappearance of such effect, we enquired about the conditions affecting the storage, maintenance and fading of the decisional representations involved in serial dependence. In experiments 2C and 2D, inter-stimulus intervals (ISI) are kept constant to investigate the effect of sequential decisions on isolation of any other aspect of the experiment. Thus, we could explore whether the maintenance of the serial dependence effect exerted by distant trials is disrupted by subsequent decisions along the same or a different feature-dimension, above and beyond any possible effect of time.

Figures 8e and 8f show the distribution of current variance reports (zRE_n) as a function of the variance in trial n-2 (only RAN trials in n-2 and n positions), split in terms of whether the intermediate trial (n-1) required a decision about variance (RAN) or mean (DIR). Trial type is pre-cued in 8e and post-cued in 8f. On visual inspection, it can be appreciated that the DIR plot presents a neat ascending slope, while this is noticeably less clear for the RAN plot. Thus, decisions about variance made two trials before are able to influence current judgments if the interposing trial required a decision about a different feature-dimension (mean), but their influence is partly disrupted if a subsequent decision about variance is interposed. Same as with the generation of serial dependence, the maintenance of its effect is not related to perceptual attention but to post-perceptual (decisional) events, since the difference persists when the required decision is post-cued.

To formally ascertain this, we selected the data where both trial n and n-2 were of RAN type and ran a Bayesian LMM on current variance report (zRE_n), with StD_{n-2}, trial n-1 type and their interaction StD_{n-2}*trial n-1 type as independent variables. The coefficient for the interaction term informed us of the effect of intermediate (n-1) trial type on

serial dependence by n-2: a positive value would indicate an increase in positive serial dependence if trial n-1 was of DIR type, compared to the reference type RAN. In Experiment 2C, this coefficient was positive, although the 95% credible intervals contained zero: B=0.0029 (-0.0017, 0.0076); likewise for Experiment 2D: B=0.0039 (-0.0008, 0.0085).

This result (albeit not statistically significant) suggests the existence of some sort of dimension-specific decisional capacity bottleneck, where only a limited number of decisional representations can be stored independently, and subsequent decisions overwrite and distort previous ones made about the same feature-dimension. This intuition could be modelled in Bayesian terms as an iteratively updated prior that is combined with new incoming information, to produce both a contextually informed judgment *and* the basis for the prior of the next trial – a process akin to a Kalman filter (Cicchini et al., 2014; Luca & Rhodes, 2016; Petzschner & Glasauer, 2011; Roach, McGraw, Whitaker, & Heron, 2017).

3. EXPERIMENT 2E: INFLUENCE OF TIME IN THE MAINTENANCE OF SERIAL DEPENDENCE

As stated above, evidence so far indicates that serial dependence in variance is driven by dimension-specific decision-making and apparently subject to dimension-specific capacity limitations affecting the decline of its effect. This echoes some sort of memory representation that is able to bias subsequent representations, as discussed by Bliss et al (Bliss et al., 2017). These authors reported that serial dependence was absent at the time of perception but increased when response was delayed, and hypothesized that judgments made about working memory representations of a percept, rather than the

percept itself, were corrupted by previous memory information. Their findings were corroborated by another team (Fritsche et al., 2017). Other authors have discussed the similarities between serial dependence and well-studied processes involving working memory limitation such as proactive interference (A. Kiyonaga et al., 2017). However this issue is far from resolved, and some authors have reported serial dependence (in position) at the time of perception (Manassi et al., 2018).

We therefore decided to examine the influence of time in serial dependence in variance, by a modification of the post-cued task-switching experiment reported above (Experiment 2D). Nevertheless, note that despite remarking the superficial similarities between serial dependence and mnemonic processes, we are still utilising a serial dependence paradigm and not an experimental design for systematic investigations on working memory.

We considered two steps at which time might have different effects:

1. 'Active' interval, i.e. time during which the current representation is actively maintained in working memory before a response is allowed (similar to Bliss' experiment (Bliss et al., 2017)). In Bayesian terms, we could see this as if the likelihood (representing probabilistic sensory information) is corrupted by noise and becomes less precise with time, so that the prior will have comparatively more weight in an optimal Bayesian combination, increasing the size of serial dependence effect (Petzschner & Glasauer, 2011). In our experiment, this active maintenance would correspond to the previously defined T1 interval between stimulus offset and post-cue onset, a time during which participants still don't know which dimension they will have to report and are supposedly maintaining a representation about both mean and variance in working memory.

2. 'Passive' interval, i.e. time between the end of a trial and the start of the next one, during which participants have no need for maintaining any information in

working memory (since each trial's stimulus and task is independent from the rest) and are (supposedly) passively waiting for the next RDK presentation. In Bayesian terms, this would correspond to a corruption of the prior by memory noise, and therefore a decrease of its influence when it is combined with the next incoming sensory information (likelihood), decreasing the size of serial dependence effect. In our experiment, this would correspond to the previously named T2 interval between response offset and next trial onset.

Therefore, in our experiment we independently manipulated T1 and T2 while keeping everything constant, and hypothesized opposite effects on serial dependence: it would increase with T1 and decrease with T2.

3.1. METHODS

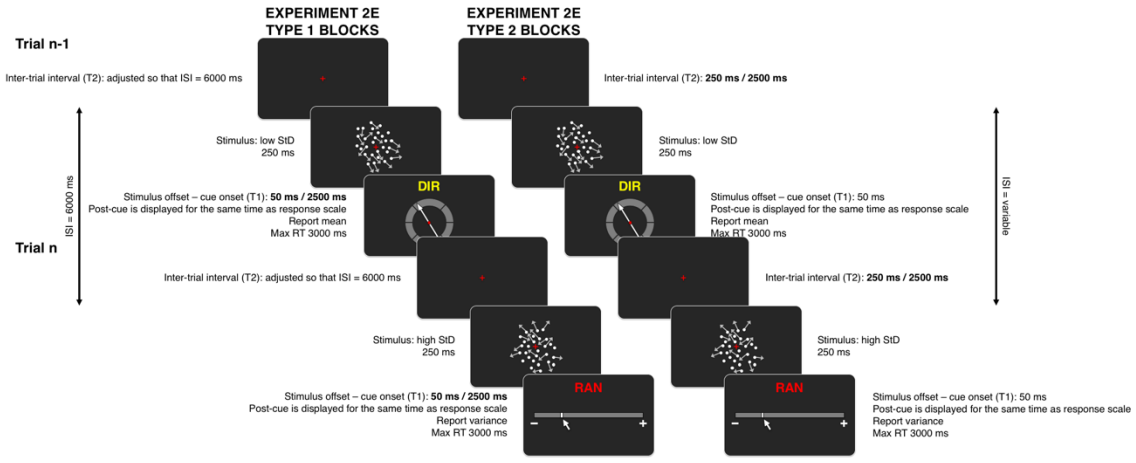


Figure 9. Experiment 2E: Methods. The experiment used a post-cued task-switch paradigm where participants had to report either the mean or the variance of an RDK. It was divided in ten blocks, of two different block types, randomly interleaved. In type 1 blocks, the interval between stimulus offset and post-cue (T1) was manipulated while keeping constant the overall interval between consecutive stimuli (ISI): half of the trials had a short (50 ms) and half

a long (2500 ms) T1. In type 2 blocks, the interval between trials (T2) -and therefore the total ISI was manipulated, while keeping T1 constant: half of the trials had a short (250 ms) and half a long (2500 ms) T2. Given previous evidence and predictions according to a Bayesian formulation for serial dependence effects, we hypothesized that a longer T1 would increase serial dependence and a longer T2 would decrease it. Abbreviations: StD – standard deviation (of the RDK motion directions), RT: response time, ms: milliseconds.

Figure 9 presents an overview of the design for experiment 2E, which was identical to the previous experiment of the series (2D: post-cued task-switching) except for the time between events. There were 10 experimental blocks (besides the practice block at the beginning) with 60 trials each; blocks were of two types, equally frequent and randomly interleaved.

In type 1 blocks we manipulated T1 interval, the 'active' interval between stimulus offset and post-cue onset. Half of the trials had 'short T1' (50 ms) and the other half had 'long T1' (2500 ms); the order of trials was pseudorandomized. The total time between stimuli (ISI) was kept constant at 6000 ms by adjusting T2 in function of T1 and the response time.

In type 2 blocks we manipulated T2 interval, i.e. the 'passive' inter-trial interval. Half of the trials had 'short T2' (250 ms) and the other half had 'long T2' (2500 ms), in a pseudorandomized order. T1 interval was always 'short' (50 ms) and the total ISI varied in function of the response time and T2.

3.2. RESULTS

Thirty-four participants (22 female, mean age 21.6 y/o) took part in the experiment, including Psychology students, members of the laboratory and payed volunteers. Two participants were excluded due to apparent erratic responses. In total, 32 participants

and 19200 trials entered the analysis. Half of them corresponded to Type 1 blocks and the other half to Type 2 blocks; in both cases, half were RAN and half DIR trials.

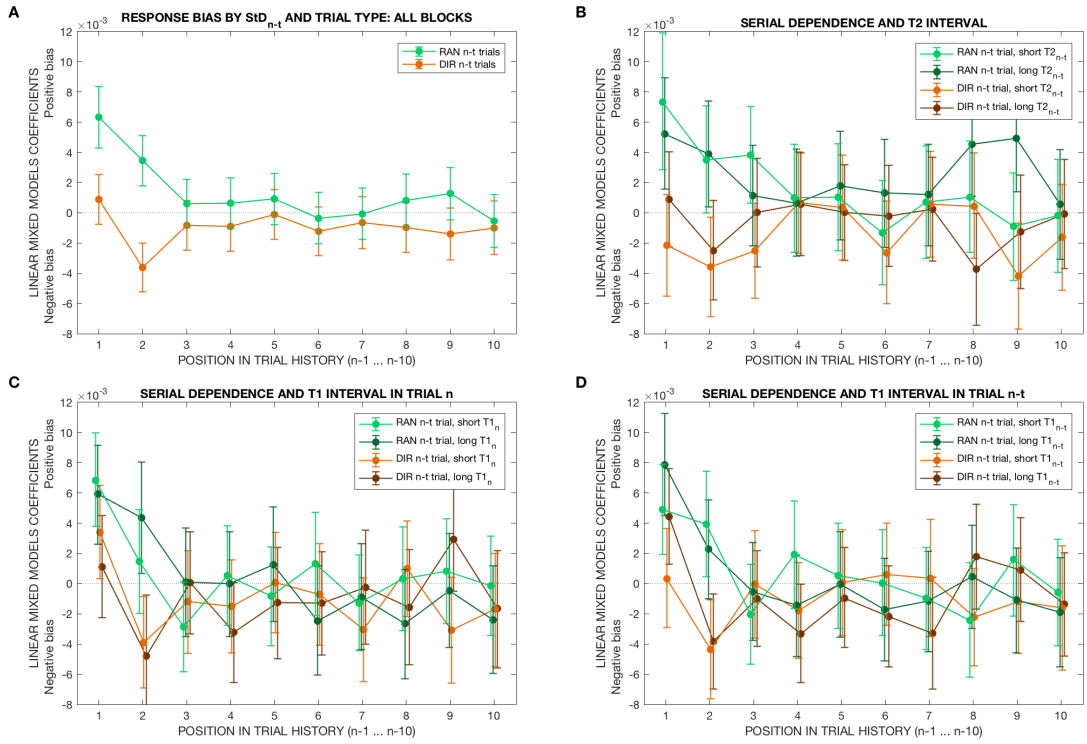


Figure 10. Experiment 2E: Results. All graphs present the fixed-effects coefficients for the influence of StD_{n-t} (t=1 … 10) on current normalized response error (zRE_n), modelled by Bayesian LMMs. The error bars represent the 95% credible intervals for the value of the coefficient. In each figure and plot, the models are applied to a specific subset of the data, in order to ascertain the influence of past decisions and time intervals in serial dependence. **10a.** Serial dependence on variance judgments (zRE_n) by StD_{n-t} and post-cued decision (RAN/DIR) presented in trial n-t (t=1 … 10). These models correspond to all the data of Experiment 2E, split by trial type (RAN/DIR), but pooling all blocks and time intervals. As in previous task-switching experiments (pre or post-cued), we observe that positive serial dependence appears only when a decision about variance was made in the considered past trial. In addition, we observe that the negative effect driven by less recent history that was found in Experiment 1 is highly reduced or even absent (for RAN trials) in this dataset, probably in relation with distinct properties of the stimulus (shorter duration) and trial structure (fewer RAN trials, longer time intervals). **10b.** Type 2 blocks: serial dependence in relation to T2 interval (time between response offset and next trial onset, when participants are passively waiting for the next trial). On visual inspection, there appears to be a trend for a weaker effect of positive serial dependence in relation with longer inter-trial intervals (T2), concerning the most recent RAN trials where positive serial dependence is observed. This may point to a decline of the effect with (passive) time; however, results are not statistically significant.

10c – 10d. Type 1 blocks: serial dependence in relation to T1 interval (time between stimulus offset and response post-cue, when participants are actively maintaining a memory representation of the mean and variance of the current trial). **10c** examines serial dependence in function of the T1 interval in the current trial (n), whereas **10d** presents the influence of manipulating T1 interval in trial n-t. **10c.** Regarding the current trial, we don't observe an increase in serial dependence by the previous (n-1) presentation in relation to longer $T1_n$, contrary to reports by other authors. However, there seems to be an enhancement of positive serial dependence by n-2 trial, although results are not statistically significant. **10d.** Regarding the previous (n-t) trial, the graphs suggest an increase of the bias exerted by n-1 StD if a longer T1 was allowed before cueing the response in that trial. However, this result is again not significant. Interestingly, positive serial dependence is observed in this case in relation to DIR n-1 trials if a long T1 interval was allowed before the cue, contrary to the absence of effect in any other task-switching dataset. This is likely explained because during the 2.5 seconds preceding the cue, participants have presumably prepared both decisions (mean and variance) about the presented RDK; thus, even if the variance decision was eventually not reported, it is still able to exert serial dependence effects on subsequent trials -similar to no-response trials in Experiment 2A.

Figure 10a presents serial dependence in variance judgments as a function of the trial type (RAN/DIR), for all block types and time intervals pooled. Each data point represents the B coefficient estimate for the effect of StD_{n-t} in zRE_n, in a Bayesian LMM. As in previous instances, we observe that positive serial dependence arises only when a decision about variance was made. The negative effect in relation to less recent history, that was clearly seen in Experiment 1, 2A and 2B, is not observed for RAN trials in the current experiment (2E) and is reduced also for DIR trials. This effect was also noticeably diminished in relation to RAN trials in Experiments 2C and 2D. As mentioned above, this may be possibly explained by the shorter duration of the RDK presentations (hence shorter exposure to the adapting stimulus), along with fewer number of interposing decisions about variance (which disrupt the competing positive serial dependence), and, additionally in the current experiment, by the enhancement of the positive effect by time manipulations, as detailed below.

3.2.1. Type 1 blocks:

Figures 10c-10d present serial dependence on variance judgments (up to trial n-10) in Type 1 blocks, as a function of the past trial type (RAN/DIR) and the T1 interval (stimulus offset – post-cue) presented in the current (10c) and the past (10d) trials.

Concerning the effect of T1 interval in the current trial ($T1_n$) we hypothesized that we would observe an increase in positive serial dependence for longer $T1_n$, similar to the reported results in Bliss et al (Bliss et al., 2017), as the current memory representation is maintained in working memory and distorted by previous information. However, we did not find such effect in relation to n-1-driven serial dependence. There seems to be an enhancement of serial dependence by n-2 trial in relation to longer $T1_n$, which would also fit our hypothesis. However, this result is not statistically significant: in a Bayesian LMM on the effect of StD_{n-2} (RAN n-2 trials only) in zRE_n, with $T1_n$ and the interaction term $StD_{n-2}*T1_n$ also included in the model, the coefficient estimate for the interaction term is B=0.0028 suggesting an increased serial dependence effect for long $T1_n$, but non-significant according to 95% credible intervals (-0.0022, 0.0077).

As for the effect of past (n-t) T1 interval in serial dependence, our predictions were less clear. Still, we deemed it would have a similar enhancing effect, since actively maintaining the n-t representation in working memory would facilitate learning and generate a more precise prior for the successive trials. Again, results were not statistically significant. There seems to be a larger effect for longer $T1_{n-t}$ when considering n-1 trial, in agreement with our hypothesis: in a Bayesian LMM for zRE_n with StD_{n-1}, $T1_{n-1}$ and the interaction $StD_{n-1}*T1_{n-1}$, the coefficient for the interaction is B= 0.0030 (-0.0014, 0.0074): positive, but not significant.

Interestingly, unlike in all other instances in our task-switching experiments, there is a clear positive serial dependence on current variance judgments by the variance of n-1 RDK (StD_{n-1}), even when this trial required a decision about mean (DIR), but *only* in those

trials wherein a long time elapsed between stimulus and response cue. For n-1 long-T1 DIR trials, the LMM coefficient for serial dependence by StD_{n-1} was B=0.0044 (0.0012, 0.0076), comparable to the effect size of serial dependence exerted by n-1 RAN trials in this and other experiments (in the current one, similar to the effect of short $T1_{n-1}$ RAN trials). In long-T1 trials, during the 2.5 seconds before the required task is cued, participants are likely preparing their decision about both possible tasks, mean and variance estimation, even if eventually only one of them will be reported. Thus, this result further supports our conclusion whereby it is the decision, and *not* the percept or the motor response, what is transferred forward to influence subsequent trials.

3.2.2. Type 2 blocks:

Figure 10b summarizes serial dependence in Type 2 blocks, split by past (n-t) trial type (RAN/DIR) and by the past T2 interval, i.e. the time between response offset in trial n-t and the onset of trial n-t+1. Our hypothesis was that a longer inter-trial interval would reduce the effect of serial dependence, by eroding the memory representation about previous decisions. Results seem to indicate a trend in this direction, with larger effect of serial dependence for shorter $T2_{n-t}$ intervals, but it was not statistically significant. In a Bayesian LMM for zRE_n with StD_{n-1} (n-1 RAN trials only), $T2_{n-1}$ and interaction StD_{n-1}*$T2_{n-1}$, the coefficient for the interaction term was B=-0.0020 (-0.0068, 0.0029). The negative term suggests that long T2 interval is associated to *less* serial dependence (as short T2 is the reference level) -however, 95% credible intervals contain zero.

In conclusion, exploration of the effect of time in serial dependence showed suggestive trends in the directions hypothesized according to a Bayesian-like, decision and memory-based conception of serial dependence: (i) an enhancement of the effect in conditions that facilitated the influence of previous memory representations on current decisions, and (ii) a decline with time in relation to a presumed fading of those memory

representations. However, these results were statistically non-significant and thus inconclusive.

CHAPTER 3: EXPERIMENT 3: INFLUENCE OF CONFIDENCE IN SERIAL DEPENDENCE IN VARIANCE

In the previous chapter we reported that positive serial dependence in our paradigm arises in relation to a dimension-specific decision made in a past trial. In light of this finding, we set out to study how confidence in such decisions modulates serial dependence in variance. Results for this experiment (Experiment 3) are reported in this chapter. We observed that confidence in the current trial decision had no influence in serial dependence, while confidence in a past trial affected the magnitude and even the sign (positive/negative) of the bias exerted by that trial's decision. We found a direct association between confidence reported in the past trial and effect size of the positive serial dependence: in other words, the attractive bias exerted by a past decision was larger the more confident had been such decision. In fact, only decisions above a certain level of confidence drove positive serial dependence effects, while negative serial dependence appeared below that level. This result was consistent with Bayesian accounts of perceptual decision-making and the interpretation of confidence as related to sensory precision (the lack of association to current confidence, however, did not fit a purely Bayesian framework). Besides, a comparison of serial dependencies in Experiments 1 and 3 provided further support to the suggestion (reported in Chapter 2) that time and capacity limitations affect positive serial dependence. The contents of this chapter have been published in the Journal of Vision, as part of the article 'Serial Dependence in the Perception of Visual Variance' (Suárez-Pinilla et al., 2018b).

In previous chapters we have examined serial dependence in visual variance, and the specific processes involved in perceptual decision-making that are responsible for its appearance. As detailed in the corresponding chapters, results of Experiments 1 and 2 indicate that positive serial dependence in variance involves mid to high-level processes, specifically decision-making about the same feature-dimension. Considering this, we questioned how confidence in those decisions modulates serial dependence. We were especially interested in the modulation of the positive (Bayesian-like) bias exerted by very recent trials, in light of Bayesian accounts of confidence as a measure of the precision of neural representations (Meyniel et al., 2015).

1. METHODS

An overview of the Experiment structure is presented in Figure 11.

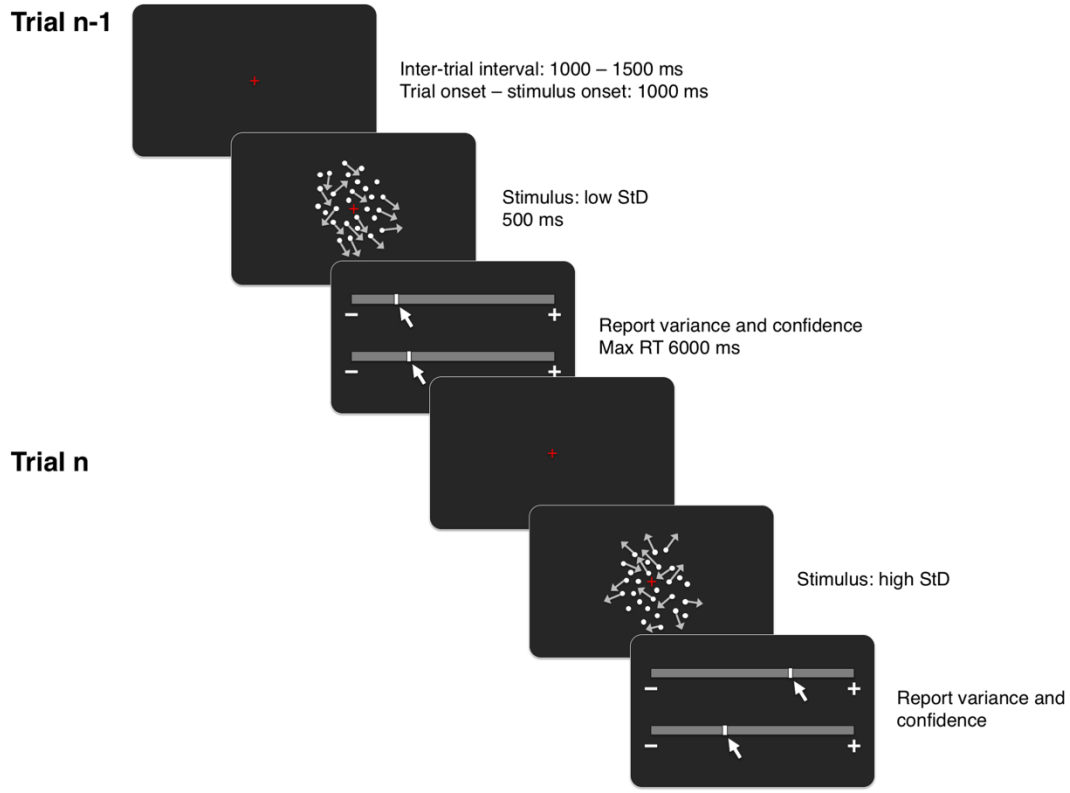


Figure 11. Experiment 3: Methods. The structure of Experiment 3 was very similar to Experiment 1, but required an additional confidence report after the variance judgment in each trial. Confidence was reported with a visual analogue scale presented below de scale for variance. Maximum response time was 6000 ms. Abbreviations: StD – standard deviation (of the RDK motion directions), RT: response time, ms: milliseconds.

1.1. STIMULI

Stimulus presentation was identical to Experiment 1; as in this case, we employed again foveal and peripheral (20º) presentations, as we considered that the interplay between decision-making, confidence and serial dependence might vary at different degrees of sensory precision.

1.2. PROCEDURE

Experiment 3, like Experiment 1, had a 72-trial practice block and eight 60-trial experimental blocks, half of which were 'foveal' and half 'peripheral'. Eye-tracking was performed in the same manner as in Experiment 1.

During the response phase of each trial, two identical visual analogue scales were displayed on the screen: the upper one for scoring 'randomness' (variance) and the lower one for confidence (see Figure 10). The initial position of each sliding bar was randomized separately, and the time allowed for responding to both items was 6 seconds. For data analysis, we obtained the numerical scores as a linear translation from the selected positions: for confidence, the score was expressed as a 0 to 1 proportion of the overall length of the line.

2. RESULTS

Twenty-two participants (17 female, mean age 19.6 y/o, standard deviation 2.42) volunteered for this experiment: all except for three members of the laboratory were first-year Psychology students. As in Experiment 1, trials without valid fixation during stimulus presentation were removed from the analysis, as well as data about trial history of valid trials involving any invalid trial. In total, 8880 trials were included in the analyses.

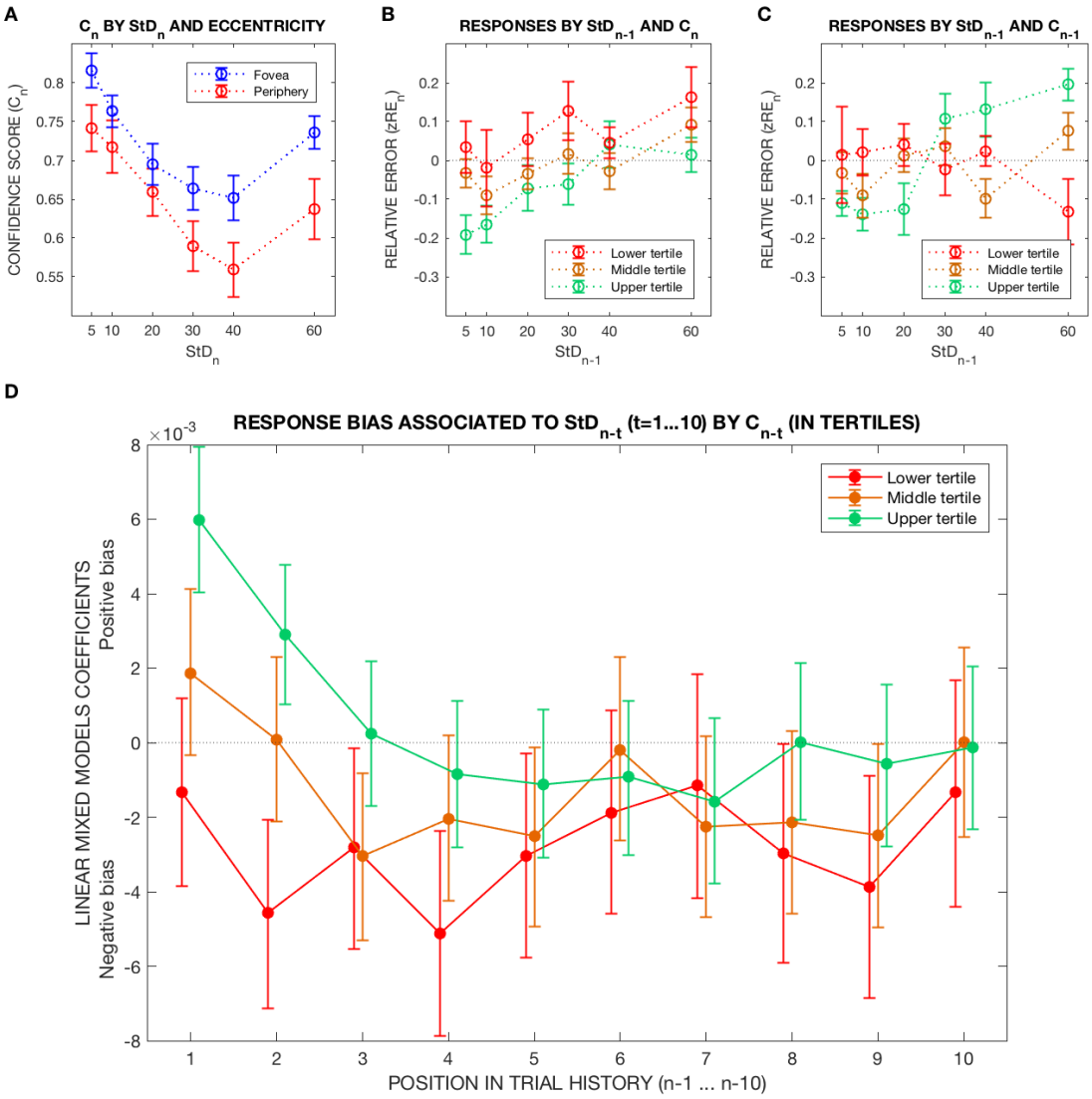


Figure 12. Experiment 3: Results. 12a. Confidence scores (C_n) by StD_n plotted separately by eccentricity. **12b – 12c.** Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by confidence reported in the current (**12b**) or previous (**12c**) trial. Confidence scores have been binned into tertiles according to each participant's distribution of reports. The error bars represent the between-subject standard error. The plots in **12b** are all ascending and roughly parallel, indicating that current confidence does not modulate serial dependence by previous trial StD. Conversely, when considering confidence reported in the previous (n-1) trial (**12c**), we observe drastically different slopes: while the high-confidence plot (upper tertile) has a clear ascending slope indicative of a positive bias, the middle-tertile plot is only mildly positive and for the lower-tertile is slightly descending, suggesting a negative bias away from low-confidence n-1 trials. **12d.** Fixed-effects coefficient estimates in 30 Bayesian LMMs with StD_{n-t} (t=1…10) as predictor of current response (zRE_n), modelled separately by confidence reported in trial n-t (C_{n-t}), binned into tertiles. The error bars represent the 95% credible intervals for the true value of the coefficient. As suggested for trial n-1 in figure **12c**, the size and direction of the bias associated with each trial position depends on the confidence reported in that position, so that the bias will be more

negative (or less positive), the lower the reported confidence, within the general trend of an increasingly negative/less positive bias as we move backwards in history.

2.1. CONFIDENCE REPORTS CORRELATE WITH THE ACCURACY AND PRECISION OF VARIANCE JUDGMENTS

Figure 12a presents the distribution of confidence scores (C_n) plotted by current stimulus StD (StD_n) and eccentricity. For both foveal and peripheral trials, a trend towards decreasing C_n for larger StD_n is observed, except for the maximal StD (60º). For each StD value, confidence scores are lower in the periphery. To test these observations, we conducted a Bayesian repeated-measures ANOVA on the effect of StD_n and eccentricity (as within-subject factors) on C_n. The best model was the one including both main effects only ($BF_{10}=6.657*10^{26}$), outperforming the full model with the interaction term StD_n*eccentricity by a factor of $BF_{\text{main effects/full}}=9.615$. This indicates that, despite the overall lower confidence scores in peripheral blocks, the relationship between different stimulus levels and confidence is the same regardless of eccentricity.

Subsequently we explored whether confidence reports were differentially shaped by response accuracy or precision – and considered also the role of eccentricity. Regarding accuracy, we defined 'error size' as the absolute value of the difference between real and reported StD: $E_n = |\ StD_n - R_n\ |$. In a Bayesian LMM with C_n as dependent variable and E_n, StD_n and their interaction as independent variables, C_n reports are inversely associated with error size (B=-0.0083 (95% credible interval -0.0103, -0.0062)) and StD_n (B= -0.0056 (-0.0071, -0.0040) and positively associated with the interaction between both (B=0.0003 (0.0002, 0.0003)). The inverse association between error size and C_n suggests that participants' reports of confidence are, at least in part, grounded in task accuracy. Furthermore, the positive sign of the coefficient estimate for the interaction term E_n *StD_n suggests that confidence tracks relative, rather than absolute error: the

inverse association between error size (defined as an absolute value) and confidence is weighted down for large StD values. When considering both error size and eccentricity, the negative association with error size remains (B_{error}=-0.0078 (-0.0102, -0.0055)), whereas foveal presentations are associated with higher confidence reports independently of task accuracy ($B_{eccentricity}$=0.0510 (0.0080, 0.0908)). However, the interaction term does not show evidence of a different evaluation of increases in error size in low compared to high eccentricities ($B_{error*eccentricity}$=-0.0013 (-0.0040, 0.0016)).

As for precision, we calculated the standard deviation of each participant's responses per StD value (σ_R) as a measure of response dispersion. Subsequently we modelled confidence by σ_R, StD and their interaction. As expected, response dispersion shows a negative correlation with confidence: B=-0.0101 (95% credible interval -0.0160, -0.0045). When adding eccentricity to this model, the main effect for σ_R is close in value (B=-0.0105 (-0.0162, -0.0050)), whereas the interaction term σ_R *eccentricity (B=-0.0003 (-0.0067, 0.0062)) suggests that the interaction between response dispersion and confidence is similar in fovea and periphery. In summary, our results indicate that confidence 'is' a measure of response precision, and, to the extent to which the latter can be considered a proxy for perceptual precision, they are in agreement with Bayesian interpretations of metacognition (Meyniel et al., 2015).

Interestingly, we observed a very strong serial dependence for confidence reports. Modelling reported confidence (by a Bayesian LMM) as a function of the report provided in the previous trial (C_{n-1}), the coefficient for the latter is B= 0.1874 (95% credible interval 0.1445, 0.2307), with an R^2=0.3188. Importantly, if we add the error size of the previous trial (E_{n-1}) to the model, as well as the interaction E_{n-1}* C_{n-1}, the coefficient estimate for C_{n-1} has a similar (even larger) value: B=0.2197 (0.1698, 0.2720). This is also the case when StD_{n-1} is included in the model, suggesting that the serial dependence in confidence scores is not only due to accuracy/attention fluctuating at timescales of several trials, nor to the direct influence of the StD in the previous stimulus, but rather

may be an expression of response inertia and/or 'confidence leak' as described in (D. Rahnev et al., 2015).

2.2. CONFIDENCE IN A PAST TRIAL DETERMINES THE DIRECTION OF SERIAL DEPENDENCE IN VARIANCE REPORTS

According to Bayesian accounts of perceptual decision-making, reliance on prior information is greater when the current sensory input is noisy or imprecise, or when the prior itself is highly precise (Cicchini et al., 2014; Petzschner & Glasauer, 2011; Summerfield & Lange, 2014). Within this framework, confidence is often regarded as a measure of the precision of the sensory signal (Meyniel et al., 2015), a consideration that is in agreement with our data. Thus, we hypothesized that high reported confidence in the current trial (C_n) would decrease any attractive pull toward previous history (with respect to variance judgments), whereas confidence in past trials (C_{n-t}) would have the opposite effect. We further reasoned that such effect of confidence in the past trials would apply mostly to very recent trials, whose information represents a more important contribution when priors are iteratively updated. Indeed, this second hypothesis is in agreement with our observation of a positive bias in variance judgments exerted only by the most recent trials (see Figure 2d as example).

Figures 12b and 12c depict zRE_n as a function of StD_{n-1}, plotted separately by current (12b) and previous (12c) trial confidence. Confidence scores have been binned into tertiles on a per-participant basis. In 12b, all three plots present an ascending, roughly parallel slope: it appears that serial dependence exerted by trial n-1 takes place independently of the confidence placed in the current judgment, contrary to our initial hypothesis. However, when we consider the influence of confidence in the previous response, we do see a striking interaction, in line with what would be expected within a Bayesian framework: low-confidence n-1 judgments do not exert any positive serial

dependence – quite the opposite, the plot has a slightly descending slope, pointing toward a negative bias in relation to StD_{n-1}. This slope is mildly ascending for medium confidence and neatly positive only for high-confidence past decisions.

In order to validate these observations, first we performed a Bayesian repeated-measures ANOVA on the effect of StD_{n-1} and C_n (confidence score in the current trial, binned into tertiles) on zRE_n. Results of a Bayesian repeated measures ANOVA are presented in Table 4a. The best model contains both main effects (StD_{n-1} and C_n) but not the interaction (BF_{10}=349.668), outperforming the model with the interaction term by a factor of $BF_{main\ effects/full}$=93.544. This provides 'very strong' evidence against the inclusion of the interaction term and indicates that confidence in the current judgment does not modulate serial dependence from the previous trial.

Subsequently we performed an analogous analysis, but with StD_{n-1} and C_{n-1} (confidence score in the previous trial, by tertiles) as within-subject factors. Table 4b presents the results of this analysis. Evidence is in favour of the null model by a large margin (31.25 times more explanatory than the second best, which includes only C_{n-1}). Nevertheless, when considering the term of interest for our hypothesis, namely the interaction $StD_{n-1}*C_{n-1}$, there is a strong evidence in favour of its inclusion compared to the model stripped of that effect (including only the two main factors): $BF_{full/main\ effects}$=26.989. Still, because none of both competing models were superior to the null model, this result must be taken with caution.

TABLE 4. Serial dependence and reported confidence - Model comparison

4a. Current trial confidence

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.002	0.010	1.000	
StDn-1	0.200	0.011	0.043	4.437	0.393
Cn	0.200	0.137	0.637	57.150	0.642
StDn-1 + Cn	0.200	0.841	21.088	349.668	1.213
StDn-1 + Cn + StDn-1 ✻ Cn	0.200	0.009	0.036	3.738	5.208

4b. Previous trial (n-1) confidence

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.923	47.768	1.000	
StDn-1	0.200	0.025	0.102	0.027	0.535
Cn-1	0.200	0.030	0.123	0.032	1.097
StDn-1 + Cn-1	0.200	8.080e-4	0.003	8.757e-4	1.439
StDn-1 + Cn-1 + StDn-1 ✻ Cn-1	0.200	0.022	0.089	0.024	1.099

Note. All models include subject.

Table 4. Experiment 3. Serial dependence (associated with trial n-1) and confidence reported in current and previous trial. Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and current (**4a**) or previous (**4b**) trial confidence. P(M): prior probability of each model, assumed to be equal for all. P(M/data): posterior probability of the model (given the data). BF_M: Bayes factor for the model. BF_{10}: Bayes factor for the alternative hypothesis (expressed by each model).

We next asked the degree to which confidence in trials located further back in history, up to n-10, influenced serial dependence of variance judgements. We split the dataset according to the confidence scores reported in each past position (C_{n-t}, discretized into tertiles within each participant's scores), and ran three Bayesian LMMs per position (30 models in total) for the association between StD_{n-t} and zRE_n at each level of past

confidence. Figure 12d presents the B coefficient estimates and 95% credible intervals for each trial position n-t (t=1…10). A marked influence of past confidence on the size and direction of serial dependence is observed, such that, when high confidence was reported in very recent trials (n-1, n-2), an attractive pull toward recent StD values is manifest, although this bias fades rapidly, being absent by trial n-3 and thereafter. Note that trials with highest confidence (upper tertile) do not exert a clear, unambiguous negative bias at any point of trial history, although some traces seem to be present from trial n-4 onwards. The largest negative bias is driven by low-confidence trials, for which it seems to appear as recently as in trial n-1 (although the credible intervals contain zero), becomes unambiguous at n-2 and peaks at trial n-4, decreasing afterwards, in contrast with the slower build-up of the negative bias seen for past trials with intermediate confidence. Thus, the reversal from positive to negative bias seen in this and previous experiments seems related to the rapid decay of the positive bias of high-confidence trials. As for the negative effect, it seems to appear as early as whenever such competing (positive) bias is not manifest, but fades more slowly than the former. Results were similar when considering foveal and peripheral blocks separately.

At first glance, the early appearance of the negative effect (after exposure to a single sub-second presentation) and its association with low confidence could suggest that it is at least in part of decisional origin, rather than exclusively a product of sensory adaptation. However, some amount of negative bias was observed in relation with past DIR trials in Experiments 2B-2D (trials in which participants were not making a decision on variance). Thus, it seems more likely that the apparent relationship between the negative effect and confidence is due to concealment of the effect in presence of the positive bias, the latter being associated with high-confidence decision-making.

On average, response times for variance reports in low, medium and high confidence trials were 1.59, 1.46 and 1.30 seconds, respectively (Bayesian RM ANOVA: BF_{10}=22288, extreme evidence for the alternative hypothesis), presumably related to subjective trial difficulty. Therefore, we sought to rule out the possibility that the effect of past

confidence on serial dependence was only related to the difference in response times, and consequently in inter-stimulus times. For each trial position up to n-10, we performed a three-way Bayesian repeated-measures ANOVA for zRE_n (as dependent variable) with three within-subject factors: StD_{n-t}, C_{n-t} (in tertiles) and $time_{n,n-t}$ (time between stimulus onset of trials n-t and n, binned in two levels with respect to the median). In all cases, the evidence for inclusion of the interaction term StD_{n-t}*$time_{n,\ n-t}$ was 'extremely' low, i.e. the Bayes factor for this specific effect was always below 1/100. This suggested that time was not confounding the reported interaction between confidence and serial dependence.

2.3. TIME AND THE ADDITIONAL CONFIDENCE REPORT MIGHT PROMOTE AN EARLIER REVERSAL TOWARD NEGATIVE SERIAL DEPENDENCE IN VARIANCE JUDGMENTS

Experiment 3 had an identical design to Experiment 1 except for the requirement of an additional report (about confidence) per trial. Consequently, an additional difference was introduced: the inter-trial time was longer in Experiment 3 that in Experiment 1 (5.06 versus 3.69 seconds, Bayesian independent samples t-test: $BF_{10}>6.690*10^7$).

Experiment 2 demonstrated that serial dependence in variance arises in relation to dimension-specific decision-making -specifically high-confidence past decisions, as shown in the current experiment. Furthermore, Experiment 2C and 2D showed that the maintenance of the serial dependence effect is disrupted by subsequent decisions along the same dimension (variance) -see Figures 8e and 8f-, and Experiment 2E suggested that time affected the size of the effect in different ways (specifically, active maintenance in working memory increased serial dependence, while passive waiting time decreased it), although results were not statistically significant. Previous work by other authors has also implicated time between successive stimuli or stimuli and response as critical contributors to serial dependence (Bliss et al., 2017; Fritsche et al.,

2017; Ryotal Kanai & Verstraten, 2005). Overall, evidence suggests that time and capacity limitations influence the effect of past on current perceptual decisions.

Considering all this, we sought to examine (post-hoc) the extent to which time and interposing decisions influenced differences in serial dependence between Experiment 1 and 3 -specifically, how they affected the decrease and eventual shift towards negative of the serial dependence effect as we move backwards in trial history.

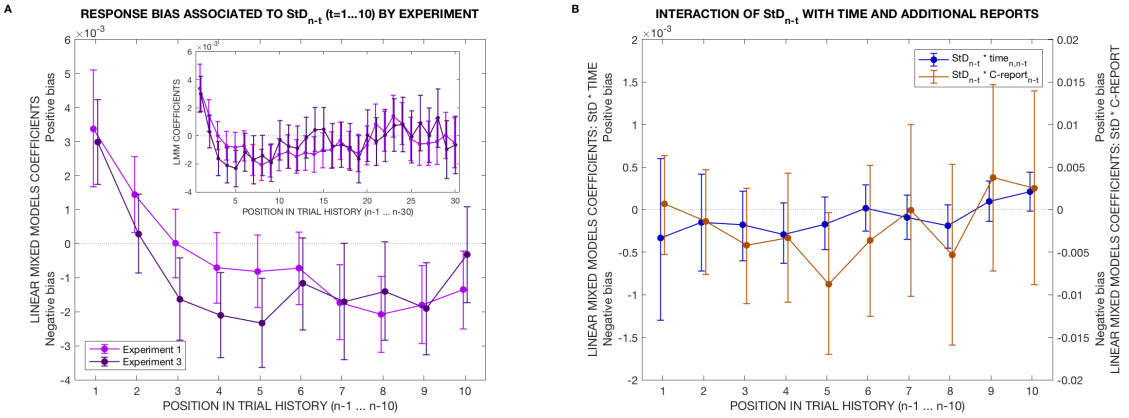


Figure 13. Comparison between Experiments 1 and 3. Both experiments have the same design except for the requirement of a confidence report (in addition to a variance report) per trial in Experiment 3. This also makes the inter-stimulus time longer, on average, for Experiment 3 compared to Experiment 1. **13a**. Fixed-effects coefficient estimates in 20 Bayesian LMMs with StD_{n-t} (t=1…10) as predictor of current response (zRE_n), with the data of Experiment 1 and 3 modelled separately. The error bars represent the 95% credible intervals for the true value of the coefficient. The shift toward negative coefficient estimates takes place at earlier trial positions in Experiment 3. The inset graph presents an extension of the analyses up to trial n-30, showing that the negative effect also fades at earlier positions in Experiment 3. **13b.** Fixed-effects coefficient estimates for the interaction terms StD_{n-t}*$time_{n,n-t}$ and StD_{n-t}*$C\text{-}report_{n-t}$ in 10 Bayesian LMMs for prediction of zRE_n, with StD_{n-t}, $time_{n,n-t}$, $C\text{-}report_{n-t}$ and all interactions as putative predictors. The variable $time_{n,n-t}$ reflects the time between onsets of the stimuli in trials n-t (t=1…10) and n. $C\text{-}report_{n-t}$ is a binary factor indicating whether confidence reports were made in all trials between n-t and n, or in none, regardless of the content of the reports (i.e. the amount of confidence). A negative interaction term with StD_{n-t} indicates a less positive/more negative serial dependence effect in relation with longer time or the requirement of an additional confidence report per trial. While credible intervals contain zero in most instances, there is a predominance of negative estimates up to n-5, which could suggest a causal role for both time and the additional confidence report in terms of promoting an earlier reversal of the bias in Experiment 3 compared to 1.

Figure 13a presents the Bayesian LMM coefficients and 95% credible intervals for the effect of StD_{n-t} (t=1…10) in current variance report as found for Experiments 1 and 3. An extension of this comparison for more distant trial positions (up to n-30) is presented in the inset graph. While the positive bias exerted by StD_{n-1} is similar in magnitude in both experiments (B=0.0034 (0.0017, 0.0051) in Experiment 1, B=0.0030 (0.0018, 0.0042) in Experiment 3), such attraction is still present at StD_{n-2} in Experiment 1 (B=0.0014, (0.0003, 0.0026)), but has virtually disappeared for Experiment 3 (B=0.0003 (-0.0009, 0.0015). Thus, in Experiment 3 the reversal to negative bias occurs as early as in trial n-3 and peaks at n-5 (B= -0.0023 (-0.0036, -0.0010)), with a similar effect size as the maximum negative bias in Experiment 1, which is seen at n-8 (B=-0.0021 (-0.0032, -0.0010)). As shown for trial positions further than n-10, negative serial dependencies also decline and disappear earlier than in Experiment 1. This earlier build-up and decline of the negative bias could be related to the longer inter-stimulus intervals in the present experiment: time might drive, hypothetically, the reversal to repulsive serial effects and its posterior fading. Results of Experiment 2B and 2D (concerning the effect of DIR trials) and low-confidence trials in Experiment 3 seem to suggest that the negative bias appears as early as whenever the conditions for the arising of a positive bias are not met. If, hypothetically, positive serial dependence declines with time, the negative effect could become evident in an earlier trial in relation to the longer inter-stimulus times observed in the present experiment. Another explanation for the earlier shift towards negative in Experiment 3 would be a disruption of the positive bias caused by the additional confidence report – especially if such Bayesian-like pull is caused by decision processes or depends upon memory to some extent. Results for Experiments 2C and 2D, concerning the effect of trial n-2 variance in relation with the decision required in trial n-1, suggested that decisions about the same feature-dimension are especially disruptive for the serial dependence effect. The extent of commonality between a primary perceptual decision and a decision about confidence is however unclear (Adler & Ma, 2016; Kepecs, Uchida, Zariwala, & Mainen, 2008; Meyniel et al., 2015; D. Rahnev et al., 2015).

To further examine the effect of time and additional decisions, we pooled all valid trials from experiments 1 and 3. To ascertain the influence of an additional decision made per trial beside the variance judgment, we defined a binary variable, named C-report$_{,n\text{-}t,}$ indicating whether or not all intermediate trials, between n and n-t, had a confidence report in addition to a variance report. Note that the content of the reports (i.e. the amount of confidence) did not affect this definition. When participants had missed at least one confidence report in the considered historical span of a certain trial, that trial was excluded from the model, in order to make the comparison unambiguous. Subsequently we built ten Bayesian LMM for zRE$_n$ (as dependent variable) in relation with three variables defined at each considered point of trial history, namely StD$_{n\text{-}t}$, time$_{n,\ n\text{-}t}$ and C-report$_{,n\text{-}t}$, and all interactions. The fixed-effects B coefficients of the interaction terms StD$_{n\text{-}t}$ * time$_{n,n\text{-}t}$ and StD$_{n\text{-}t}$ * C-report$_{,n\text{-}t}$ are plotted in figure 11b, for trials n-1 to n-10 as predictors of current variance judgment. A negative interaction coefficient would indicate a comparatively less positive/more negative serial dependence effect at that position in relation to longer time or the extra report, respectively. At all positions, credible intervals for both interaction terms contain zero (except for StD$_{n\text{-}t}$ * C-report$_{,n\text{-}t}$ at n-5). However, there is a predominance of negative values for both interaction terms within the recent half of the considered span of trial history, up to trial n-5. Thus, although results are inconclusive regarding the causes for the different patterns of serial dependence in Experiments 1 and 3, the 'mostly-negative' interaction terms StD$_{n\text{-}t}$ * time$_{n,n\text{-}t}$ and StD$_{n\text{-}t}$ * C-report$_{,n\text{-}t}$ suggest that both time and the additional confidence report might promote a less positive / more negative serial dependence in variance and thus contribute to the observed earlier reversal in the direction of the bias. Although these results are non-significant and based on post-doc analyses, they are in line with results of Experiment 2, also suggesting a decision capacity bottleneck as well as a memory decay.

CHAPTER 4: EXPERIMENT 4: SERIAL DEPENDENCE AND PERCEPTUAL AWARENESS

In previous chapters we have described evidence for serial dependence in variance as a process arising from confident decision-making, rather than low-level perception. Here we present a modification of the well-studied paradigm of serial dependence in orientation which further confirms that, even for a perceptual dimension mainly processed in the primary visual cortex (V1), serial dependence cannot arise only by local activity changes in sensory areas, unlike negative adaptation. We blocked perceptual awareness in half of the trials by continuous flash suppression (CFS), a method that is known to preserve low-level sensory after-effects in relation to unseen stimuli. However, positive serial dependence did disappear under these conditions, showing that it depends on processes beyond early visual areas, different from those responsible for the encoding of sensory information about the stimulus.

Our experiments so far indicate that serial dependence in variance is independent of low-level properties of the stimulus (such as spatial location) and driven by high-confidence, dimension-specific decisions that exert a Bayesian-like influence on subsequent judgments. This seems at odds with some studies on serial dependence that propose that it is at least partially driven by sensory processes (Cicchini et al., 2017; Fischer & Whitney, 2014; John-Saaltink et al., 2016; Manassi et al., 2018). These studies have typically examined serial dependence in low-level feature-dimensions (orientation, spatial position) that are processed by early cortical areas of the visual hierarchy. Although the locus of variance processing is not completely defined, it is likely that the

line between perceptual and post-perceptual processes is more blurred for statistical properties (Payzan-LeNestour et al., 2016; Storrs, 2015).

Considering all this, we decided to dissociate low-level sensory processing from higher-level mechanisms in a well-studied dimension such as orientation, by employing a continuous flash suppression (CFS) paradigm. CFS is a method that exploits binocular rivalry by presenting a flashing high-contrast mask in the dominant eye in order to effectively and continuously block awareness of an alternative image presented in the non-dominant eye, even for long periods of time (Tsuchiya & Koch, 2005). CFS has been employed to explore the differences between unconscious and conscious processing, and thus to gain insight on the involvement of the different levels of the visual system in the construction of visual awareness. Several studies have established that adaptation to low-level sensory information is attenuated, but not suppressed, under CFS and other instances of interocular suppression (Blake, Tadin, Sobel, Raissian, & Chong, 2006; Lin & He, 2009; Maruya, Watanabe, & Watanabe, 2008), indicating that sensory processing in early visual areas is preserved (albeit not unaltered) in absence of awareness. Interestingly, numerosity adaptation (Liu, Zhang, Zhao, Liu, & Li, 2013) and processing along the dorsal stream of the visual system (Almeida, Mahon, Nakayama, & Caramazza, 2008) also seem to be preserved, while high-level adaptation after-effects involving the ventral stream are absent in CFS (Stein & Sterzer, 2011).

We reasoned that, if serial dependence arises in relation to sensory processes, for example as a result of exposure-dependent changes in the gain or tuning of sensory neurons (as proposed in (Fischer & Whitney, 2014)), it may exhibit a similar relationship with CFS than other exposure-dependent sensory processes such as negative adaptation. Concerning serial dependence in a low-level dimension such as orientation (for which the locus of adaptation after-effects is in the primary visual cortex, V1 (Hubel & Wiesel, 1977)), if serial dependence is sensory-driven it may be preserved, though attenuated, in a CFS experiment. Thus, for this experiment we used serial dependence

in orientation, and not variance, because predictions for the locus of sensory processing and the effect of interocular suppression were more straightforward.

In the present experiment (named Experiment 4), participants looked at a dichoptic presentation while using a stereoscope. A sequence of Gabor patches was presented to their non-dominant eye while a mask suitable for CFS was displayed to their dominant eye in only half of the cases, rendering half of the Gabors (theoretically) invisible. A report about the Gabor orientation was required in each trial. We examined serial dependence in relation to the orientation of visible (unmasked) compared to invisible (masked) Gabors in trial history, aiming to shed light on the potential involvement of low-level sensory processes in serial dependence.

As a secondary question, we considered the possible relationship between regression effects (i.e. contextual modulation by global statistics) and serial dependence (contextual modulation by recent stimuli). As mentioned in the General Introduction (Part I, section 3.1), Bayesian models wherein the prior is built up as successive trials are supplied to the observer are able to explain global and recency biases as two aspects of the same phenomenon (Kalm & Norris, 2017; Petzschner & Glasauer, 2011). In light of this, we dissociated the global statistics of masked and unmasked Gabors in order to explore whether or not CFS masking influenced regression effects in the same manner as serial dependence.

1. METHODS

Figure 14 presents an overview of the experimental structure.

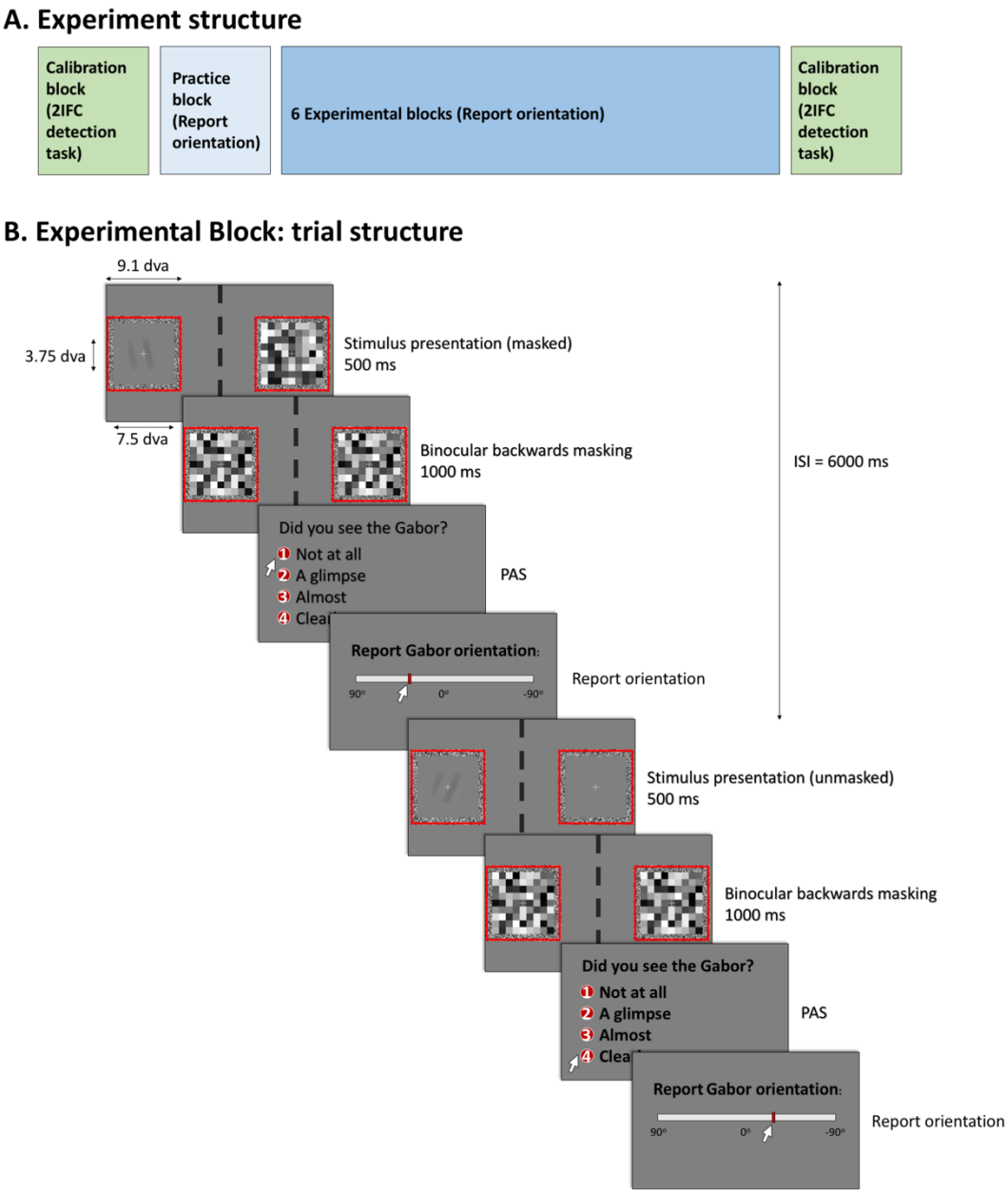


Figure 14. Experiment 4: Methods. 14a. Experiment structure. There were two calibration blocks at the beginning and end of the experiment, where Gabor contrast suitable for effective CFS masking was determined through a 2IFC detection task. After the first calibration block, there was a practice block and 6 experimental blocks which required reporting the orientation of a Gabor patch. Half of the trials were masked and half were not: data analysis ascertained serial dependence in orientation in relation to unmasked compared to masked Gabors. **14b.** Trial structure in an experimental block. A dichoptic setting displayed a Gabor patch in the non-dominant eye (except in catch trials, see main text); in the dominant eye, a CFS mask was presented in half of the trials (masked trials) and an empty

background in the other half (unmasked trials). Binocular backward masking was presented for 1000 ms in all trials. Participants had to score Gabor visibility by the perceptual awareness scale (PAS) and report the orientation of the Gabor with a visual analogue scale.

1.1. APPARATUS

The experiment was programmed in MATLAB 2017b (MathWorks Inc., Natick, US-MA) with Psychtoolbox 3.0.14 and displayed on a LaCie Electron 22BLUE II 22'' with screen resolution of 1024 x 768 pixels and refresh rate of 100 Hz. A custom-made stereoscope was deployed for enabling a dichoptic presentation; participants' head position was stabilized with a chin and forehead rest, with their eyes aligned at the height of the screen centre at 50 cm distance.

1.2. DICHOPTIC PRESENTATION

As shown in Figure 14b, a dichoptic presentation was displayed on the screen, with the two fields contained on two square red frames of 9.1 dva side length. A fusion contour of 0.8 dva width was contained within each frame, formed by randomly distributed black and white pixels to promote stable binocular alignment. This contour left an inner grey square area of 7.5 dva side (9.1 – 2*0.8) within each frame, where the stimulus or the mask was contained. In the centre of this area, a white fixation cross was displayed, spanning 0.8 dva horizontally and vertically, with a line thickness of 0.04 dva. Initially, both frames were horizontally aligned on the screen, at a horizontal distance of 18 dva between frame centres, or 9 dva to the left and right of the screen centre. However, the participant could adjust the position of the left frame, horizontally and vertically, using the mouse at the beginning of each experimental block, to achieve a perfect binocular overlap between displays.

1.2.1. Stimuli

The stimulus was a Gabor patch presented in the centre of the display frame for the non-dominant eye (therefore at 0 dva with respect to participant's gaze position), at a peak contrast that was determined for each participant at the beginning of the experiment, and adapted on a trial-by-trial basis to ensure that the Gabor remained invisible under the CFS mask throughout the experiment – see 'Procedure' section below for further detail. The spatial frequency of the Gabor was 0.5 cycles/dva and the width (standard deviation) of the Gaussian envelope was 1.5 dva. The Gabor was contained within an invisible square window (concentric with the frame of the display for the non-dominant eye) of 3.75 dva side -i.e. half the length of the inner area of the display frame, after discounting the fusion contour. This limitation to the extension of the Gabor was set to further ensure that the CFS mask continuously overlapped the stimulus, even in presence of a small transient binocular misalignment. The Gabor was presented for 500 ms.

The Gabor patch would present one of the following orientations: -60º, -48º, -36º, -24º, -12º, 0º, 12º, 24º, 36º, 48º, 60º, representing angular distance to vertical (negative sign for clockwise tilt, positive for counter-clockwise). See 'Procedure' below, regarding the presentation of the different orientations throughout the trial sequence.

1.2.2. Mask

The CFS mask was displayed within the frame corresponding to the dominant eye, and consisted on a 9x9 checkerboard pattern spanning 7.5 dva horizontally and vertically, formed by squares that took shades of grey extracted from a Gaussian distribution with mean 128 (0-255) and standard deviation 255 (bounded at 0 and 255). The shades of each component square flashed at 10 Hz. The mask was presented for the duration of the Gabor patch (500 ms).

Backward masking of the same properties was presented binocularly for 1000 ms after stimulus and CFS mask offset, or after stimulus offset in unmasked trials (see 'Procedure' below). In this case, the greyscale intensity of the component squares was the same for the display presented in both eyes, also flashing at 10 Hz.

1.3. PROCEDURE

The experiment lasted for about 90 minutes and was formed by a calibration block at the beginning, a training block, six experimental blocks and another calibration block at the end.

1.3.1. Calibration blocks

At the beginning of the experimental session, eye dominance was determined by the Porta test (Wade, 1998). After this, a calibration block was run to determine the initial contrast of the Gabor, ensuring the effectivity of the CFS mask. The calibration block was a two-interval forced-choice (2IFC) task wherein two CFS masks were presented in the dominant eye for 500 ms each, separated by a 1500 ms interval. A Gabor patch was presented in the non-dominant eye at the same time as one of the masks - 50% of the times under each. Participants had to report whether the Gabor had been presented under the first or the second mask. The calibration block had 96 trials and used 8 different peak Michelson contrast levels for the Gabors, each repeated 12 times throughout the block: 0.01, 0.05, 0.075, 0.1, 0.15, 0.25, 0.50, 1. A cumulative Gaussian Psychometric curve was fitted on participants' responses and the contrast level corresponding to 55% correct responses (guessing rate: 50%) was set as the initial contrast of the Gabors presented later on in the practice and experimental blocks. If

curve fitting was unsuccessful or if the output contrast value was too low or high (below 0.05 or above 0.75), the initial contrast was set at 0.35.

At the end of the experiment (after the six experimental blocks) a second calibration block was run, to rule out that adaptation to the CFS mask throughout the experiment had rendered the employed contrast value(s) visible.

1.3.2. Experimental blocks

After the first calibration block there was a practice and six experimental blocks. Unlike the calibration block, which was a 2IFC detection task, the main experimental blocks required reporting the Gabor's orientation in each trial. In the practice block feedback was provided at the end of each trial, by showing the participant's response along with the veridical Gabor orientation in two separate response scales -see below for the response procedure.

Each block had 90 trials, of which, in 84 trials, a Gabor was presented in the non-dominant eye: half of them (42 trials) were 'unmasked trials', where no mask but only the grey background was presented in the dominant eye, whereas the other half, randomly interleaved with the rest, were 'masked trials', where a CFS mask was displayed. The remaining six, also interleaved, were 'catch trials', where the CFS mask was presented in the dominant eye but there was no Gabor in the other display frame.

As described in the 'Stimuli' section, the Gabors could take 12 possible orientations: -60º, -48º, -36º, -24º, -12º, 0º, 12º, 24º, 36º, 48º, 60º (negative sign indicates clockwise with respect to vertical, and positive sign counter-clockwise). These values were organized into two distributions: D1, ranging from -60º to 12º, and D2, from -12º to 60º. In half of the blocks, unmasked Gabors would take orientations from D1 (each value would be repeated 6 times through the block), and masked Gabors from D2; in the other

three blocks, randomly interleaved, it would be the opposite way. This meant that the three values present in both distributions (-12º, 0º, 12º) were twice as frequent (6x2=12 repetitions each) compared to the others. Thus, the global mean orientation of all Gabors of the block was 0º, while the mean orientation of D1 and D2 was -24º and 24º, respectively. In other words, in half of the blocks the mean orientation of the unmasked Gabors was -24º (and 24º for the masked trials); while in the other half the unmasked Gabors would have a mean orientation of 24º. We applied this design to explore whether regression-to-the-mean effects, i.e. the influence of global statistics (Hollingworth, 1910), was exerted by all orientations or only the ones that reached awareness -in parallel with the exploration of the relationship of visual awareness and the effect of recent history (serial dependence), which was the key question in our study.

Experimental blocks: trial structure

As depicted in Figure 14b, the structure of a trial in experimental blocks was as follows: a test Gabor patch was displayed immediately after trial onset during 500 ms, in the non-dominant eye, whereas either an empty background or a CFS mask were presented for the same time in the dominant eye, depending on whether the trial was 'unmasked' or 'masked'. The exception were the catch trials, where an empty background and a CFS mask were presented for the same time in the non-dominant and dominant field, respectively. Immediately after this, a binocular backward mask was presented for 1000 ms in all trials.

Subsequently, the participant had to report the visibility of the Gabor patch according to the perceptual awareness scale (PAS) (Ramsøy & Overgaard, 2004). To the question (displayed on the screen) 'Did you see the Gabor?', they had to select the appropriate score with the mouse: 1- Not at all, 2- A glimpse, 3- Almost sure, 4- Clearly saw it. The desirable and usual outcome for our experiment was that participants scored 1 for masked trials and 2 or more (usually 4) for unmasked trials. As described above, the

initial contrast of the Gabors was determined during the first calibration block. However, if the participant scored anything other than 1 in a masked trial, the Gabor contrast was lowered by 0.05 for all the following trials. At the beginning of a new experimental block, the initial contrast was the final value employed at the end of the previous block. This was done to ensure that calibration had effectively rendered the masked Gabors invisible and that they remained invisible during the entire experiment despite any adaptation to the mask.

After responding to the PAS, participants had to report the orientation of the trial Gabor by adjusting the position of a sliding bar along a horizontal line which ranged between 90^{o} and -90^{o}. We chose not to use a rotating bar in order to preclude tilt after-effects (or positive serial dependences) driven by the response scale itself. Instead, participants were trained throughout the practice block to translate angular orientation into a linear scale. The initial position of the slider was randomized for each trial. The maximum response time for both PAS and orientation report was 3500 ms, and the time between consecutive trial onsets (inter-stimulus interval, ISI, or inter-mask interval if a catch trial is involved) was kept constant at 5000 ms by adjusting the interval after response offset in function of the response time.

1.4. STATISTICAL ANALYSIS

Statistical analyses (detailed in the Results section) were performed on Matlab 2016a (MathWorks Inc., Natick, US-MA), R 3.4.2 (The R Foundation for Statistical Computing, http://www.R-project.org) and JASP (JASP Team (2017). JASP (Version 0.8.3.1, Mac OS X – El Capitan (10.11)).

2. RESULTS

2.1. PARTICIPANTS AND DATA SELECTION

Eighteen participants volunteered to the experiment (12 female, mean age 26.2 y/o) in exchange for £10 payment; two participants did the experiment twice. The entire sample comprised 10800 experimental trials.

In order to make Gabor visibility unambiguous according to self-report, we removed all unmasked trials in which participants had reported a PAS score of 1 (15.96% of unmasked trials) and all masked trials that had been scored 2 or higher according to PAS (10.59%, with 8.88% scored as 2 and 0.80% as 4). We also removed all information corresponding to the excluded trials out of the history of other trials: for example, if a trial n had in position n-1 another trial that had not been scored according to its category (an unmasked trial scored with a PAS of 1 or a masked trial scored higher than 1), this trial n was considered to have no valid data for analyses regarding n-1 history. Concerning catch trials, they were scored 2 or higher 8.12% of times (7.28% scored as 2 and 0.28% as 4).

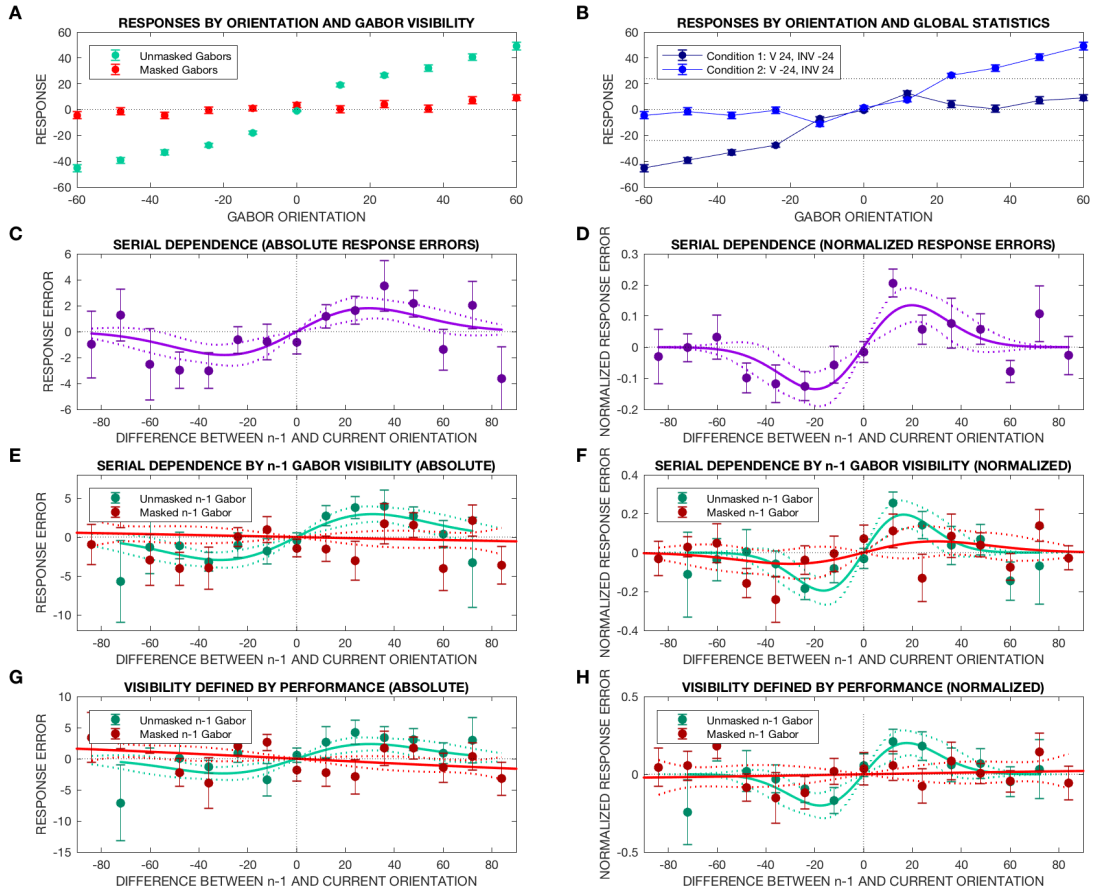


Figure 15. Experiment 4: Results. 15a. Response by current Gabor visibility (unmasked/masked). **15b.** Response by global statistics. In Condition 1 blocks, the mean orientation of the visible (unmasked) Gabors was 24º, and the mean orientation of the invisible (masked) Gabors was -24º. In Condition 2 blocks it was the opposite way. If there were regression to the mean effects exerted by visible Gabors only, we should observe comparatively more positive (more counter-clockwise) reports in Condition 1, especially for the three orientations that are presented with equal frequency in unmasked and masked Gabors across both conditions: -12º, 0º, 12º. However, we do not find any trace of effect in that direction. **15c – 15d.** Serial dependence by previous (n-1) Gabor orientation. Both figures show the distribution of response errors (absolute in **15c**, normalized within participant and Gabor orientation in **15d**) as a function of the difference between previous and current Gabor orientation, $\Delta OR_{n-1,n}$. A positive $\Delta OR_{n-1,n}$ indicates that the previous Gabor was tilted more XCW (counter-clockwise) than the current one, and vice versa. Likewise, a positive response error indicates that the report was more XCW than the current veridical orientation (**15c**), or than the average report provided by that participant for that orientation (**15d**). Responses correspond to unmasked (current, n) Gabors, while the previous (n-1) stimulus could be either unmasked or masked. The error bars represent the average reports and between-participant standard errors. The continuous line represents the best-fitting derivative of Gaussian function to reflect the non-linearity of the serial dependence effect; the dotted lines present the 95%

prediction intervals for the true value of the function, computed non-simultaneously per $\Delta OR_{n-1,n}$ value. A neatly positive serial dependence effect is observed in relation to n-1 Gabor orientation, of similar tuning width and effect size to those reported in previous experiments. **15e – 15f.** Serial dependence by previous (n-1) Gabor orientation, plotted separately by previous (n-1) Gabor visibility: unmasked-visible (green plots), masked-invisible (red plots). **15e** presents the absolute deviation from veridical, while **15f** presents response errors in normalized form. Apart from the classification by n-1 visibility, data and fitted curves presentation is identical to that described for **15c- 15d**. We observe that serial dependence only arises in presence of visual awareness of the previous Gabor. Thus, low-level sensory processing of orientation is insufficient to cause a serial dependence effect on subsequent presentations. **15g-15h** present the same results that **15e** and **15f**, respectively, but removing those blocks wherein there existed *any* evidence for a positive correlation between masked, reportedly *unseen* Gabors and orientation reports, as well as those blocks where there was *no* significant evidence for a positive correlation between unmasked, reportedly *seen* Gabors and orientation reports. In other words, we applied a (highly strict) objective criterion for data selection based on performance; note, however, that, while it is possible that above-chance performance for 'unseen' trials in some sessions may indicate that the participant' report of invisibility was not entirely accurate, it may as well indicate that above-chance performance is still possible for unseen orientation stimuli -an question that is outside the scope of this thesis. Therefore we are *not* claiming that this strict data selection is a better reflection of serial dependence by Gabor visibility than figures **15e** and **15f**. Nevertheless, the distinction becomes even more clear here: there is no serial dependence in relation to n-1 masked Gabors (for which orientation reports were at chance).

2.2. OVERALL RESULTS BY VISIBILITY

Figure 15a presents the distribution of responses by orientation in visible and invisible trials. We performed a Bayesian RM ANOVA on orientation response with two within-participant factors: Gabor orientation and masking. According to evidence from this analysis, the most explanatory model contained both main effects (Gabor orientation, masking) and their interaction, indicating that, as is evident on visual inspection, the condition of masking affected the relationship between stimuli and responses ($BF_{10}=1.659*10^{144}$).

We also examined this stimulus-response relationship separately for masked, reportedly unseen Gabors (with a PAS score of 1), by means of a Bayesian RM ANOVA on response, with Gabor orientation as within-participant factor. Strikingly, we did find a statistical association between stimulus and response within this subset of Gabors: Bayesian RM

ANOVA on response by orientation in masked trials: BF_{10}=68.376, indicating very strong evidence in its favour compared to the null model. Indeed, a very mild ascending slope can be discerned on visual inspection of the plot (Figure 15a). This finding prompted detailed analysis on individual datasets. Examining the trial-by-trial correlation between Gabor orientation and response, in masked trials wherein the participant reported not to see the Gabor at all (PAS=1), we found that for most participants, evidence opposed any trace of correlation (BF_{+0}<1/3, analysing all experimental blocks either pooled or separately). However, there were two participants with at least moderate evidence in favour of a positive correlation (BF_{+0}>3), and other three with anecdotal evidence (BF_{+0}>1). When considering the six experimental blocks separately, we found that a statistical association between masked stimulus and response was encountered in some specific blocks: in total there were seven blocks wherein evidence for a positive correlation was at least moderate (BF_{+0}>3), and other seven with anecdotal evidence (BF_{+0}>1), out of the 120 blocks of the entire sample. These findings might have two alternative explanations: either some participants exhibit above-chance performance in an orientation estimation task in absence of awareness (i.e. some sort of *blindsight (Weiskrantz, Warrington, Sanders, & Marshall, 1974)*), or their reported lack of awareness of the masked Gabor (PAS=1) was not as complete as claimed. The fact that the stimulus-response correlation is found for only a minority of participants makes us consider the second option as the most likely. Furthermore, the finding of such correlation in some blocks and not others from the same participant might be explained by some misalignment in the dichoptic presentations that the participant adjusted at the beginning of each block, so that the overlap between mask and Gabor might have not been perfect. However, in absence of proof for these conjectures, we cannot dismiss the PAS report, according to which all masked Gabors that entered our analyses were unseen.

To further examine the pattern of responses to masked trials, we compared the shift in the position of the slider along the response scale in unmasked compared to masked trials (the absolute distance between initial position and actual response). In a Bayesian

t-test, evidence was neither for nor against the existence of a difference between both trial types: mean shift 51.00º (SE 0.887), 41.77º (SE 4.7722) for unmasked and masked Gabors, respectively, Bayesian paired-samples t-test: BF_{10}=0.973. This indicated that most participants were not merely reporting the 'default', randomized initial position of the slider in masked trials (which would have rendered a shift of 0º), although there were important differences between participants, as suggested by the large standard error for the shift of the slider in masked trials.

2.3. OVERALL RESULTS BY GLOBAL STATISTICS: NO OBSERVED INFLUENCE OF VISUAL AWARENESS IN REGRESSION EFFECTS

Figure 15b shows the distribution of responses by orientation and condition: in condition 1, the mean orientation of unmasked and masked Gabors is -24º and 24º, respectively; in condition 2 it is the opposite. As suggested by visual inspection of the plots, we found no evidence for an effect of the global statistics of visible Gabors on responses. To formally ascertain this, we selected the orientations common to both distributions, D1 and D2 (see 'Methods: Experimental blocks' section), i.e., -12º, 0º and 12º, and ran a Bayesian ANCOVA with response as dependent variable, orientation as covariate, condition as fixed factor and participant's ID as random factor. There was strong evidence against any effect on 'condition' on responses by orientation, in all trials pooled as well as when analysing responses to unmasked trials separately (BF_{10}<0.100 in both cases); when selecting masked trials only, evidence was neither for nor against an effect of 'condition' (BF_{10}=0.946).

In summary, according to our analyses, influence of global statistics on responses (regression effect) was not mediated by Gabor visibility. This may suggest that high level processing is not necessary to generate a global prior. Nevertheless, it is also possible that the absence of observable effect was due to our experimental design. It is very likely

that the length of the blocks (84 Gabors, 42 visible) was insufficient to 'learn' a prior for the global statistics (at least until very late in the trial sequence) -in fact, it may even be the case that during the first part of the block participants exhibited regression effects toward the condition learnt in the previous block. Another possible issue is that the linear response scale, centred on the global mean (0^o), might have produced a strong regression effect toward 0^o that prevailed over other global effects.

2.4. SERIAL DEPENDENCE IN ORIENTATION: REPORTS ARE ATTRACTED TO N-1 GABOR ORIENTATION

For ascertaining serial dependence, we selected current trials with unmasked (visible) Gabors and analysed whether the reported orientation of the stimulus was influenced by previous history. Previous studies on serial dependence in orientation have described an attractive bias exerted by previous stimuli, whose effect size depends on the similarity between previous and current stimulus magnitudes (Cicchini et al., 2017; Fischer & Whitney, 2014; Fritsche et al., 2017). This relationship is similar, but of opposite sign, to a well-known property of negative adaptation (for example, the tilt after-effect, TAE (Gibson & Radner, 1937; Magnussen & Kurtenbach, 1980)). In both cases, the size of the effect is maximal when the difference between previous and current stimulus is relatively small, and declines or disappears for larger differences (or even reverts for very large differences (Fritsche et al., 2017; Gibson & Radner, 1937; Wenderoth & Johnstone, 1988)). In our experiments on variance, we also found such relationship with stimulus similarity, although results were compromised due to floor and ceiling effects given the particularities of the experimental design (see Chapter 1, section 2.2.2).

Considering all this, we explored serial dependence in our data by analysing response errors in function of stimulus dissimilarity. Response errors were calculated as $RE_n=R_n-$

OR_n, where R_n is the current response and OR_n is the orientation of the Gabor in the current trial. In our experiment, convention has it that clockwise (CW) orientations have negative sign and counter-clockwise (XCW) orientations have a positive sign. Therefore, a positive RE_n indicates that the reported orientation was more XCW than veridical, and a negative RE_n indicates the that the report was more CW.

Figure 15c presents the average RE_n (and between-participants standard error) plotted by orientation difference between previous and current Gabor: $\Delta OR_{n-1,n}=OR_{n-1}-OR_n$. The current Gabor is always unmasked, but the previous trial could be unmasked or masked -all cases are pooled together in Figure 15c. A positive $\Delta OR_{n-1,n}$ indicates that the previous Gabor orientation was more XCW compared to the current one, while a negative $\Delta OR_{n-1,n}$ indicates that the previous one was tilted more CW. Therefore, an attractive bias (a pull exerted on the current response by the previous trial Gabor) would be represented by RE_n reports, on average, of the same sign as $\Delta OR_{n-1,n}$: a more XCW Gabor in recent history would determine a XCW bias on orientation estimation, and vice versa. A repulsive bias (such as negative adaptation effects) would translate into RE_n of opposite sign as $\Delta OR_{n-1,n}$.

As mentioned before (Chapter 1, section 2.2.2), the non-linear relationship between stimulus (dis)similarity and effect size of the induced bias has been characterized by a derivative of Gaussian function (DoG), both for negative adaptation (Heron et al., 2012) and (positive) serial dependence (Fischer & Whitney, 2014; Fritsche et al., 2017). We therefore followed previous studies and fitted our data to a DoG function of the shape:

(1)

$$y = \frac{ax}{\sqrt{2\pi}\sigma^3} e^{\frac{-x^2}{2\sigma^2}}$$

where *x* is the difference between previous and current stimulus ($\Delta OR_{n-1,n}$), *y* is the current normalized response error (zRE_n), and *a* and σ are the coefficients to be fitted,

representing the amplitude and the width (standard deviation) of the function, respectively. Fitting was achieved by non-linear least-squares method.

The best-fitting function is shown in Figure 15c along with the experimental data and the 95% prediction intervals for the value of the function computed non-simultaneously at each $\Delta OR_{n-1,n}$. The coefficient estimates and 95% confidence intervals for the function are: a=6477 (1586 – 11370), σ=29.48^o (18.28 – 40.67). In other words, the amplitude of the serial dependence effect exerted by trial n-1 was maximal when the difference between the orientation of the previous and current trial was 29.48^o. For that value of $\Delta OR_{n-1,n}$, the bias on the current response was 1.80^o toward the previous trial orientation -an attractive bias, as indicated by the positive sign of the amplitude a coefficient. The fact that the 95% confidence intervals for a did not contain zero indicates that the effect was statistically significant. This is also illustrated on Figure 15c by the 95% prediction intervals for the value of the function, which show a clear attractive effect even at the prediction boundary that is closest to the horizontal axis. The width of the function (σ) is of similar magnitude as that reported by Fischer and Whitney (27.78^o) (Fischer & Whitney, 2014); however the amplitude of the effect is much lower in our case (1.80^o versus 8.19^o). This could be related to our employment of monocular stimuli, half of them masked by CFS; nevertheless, effect sizes reported by Fritsche and colleagues are similar to our own -in fact slightly lower (1.15^o), but more narrowly tuned (σ=17^o) (Fritsche et al., 2017).

In order to remove the influence of systematic individual biases, we repeated the analysis with response errors normalized within each participant and current orientation (zRE_n). A positive zRE_n indicates that the report was more XCW, comparatively, than other reports given by the same participant for the same (unmasked) Gabor orientation; a negative zRE_n indicates a comparatively more CW estimate, compared with others given by the same participant for the same stimulus. Thus, normalization explores response biases by attending only to trial-by-trial variability within participant and stimulus. As with absolute response errors, an

attractive bias by previous orientation would be reflected by $\Delta OR_{n-1,n}$ and zRE_n of the same sign: a more CW Gabor in trial n-1 would produce a more CW report for trial n Gabor, compared with other reports provided by the same individual and for the same veridical orientation, and vice versa.

Figure 15d shows the distribution of average zRE_n (and between-participant standard errors) as a function of $\Delta OR_{n-1,n}$. The best-fitting DoG function and 95% prediction intervals are also depicted. Its coefficient estimates and 95% confidence intervals are: a=218 (75.15 – 360.8), σ=19.78^o (13.29 – 26.27), representing a peak attractive effect of 0.1348 z-scores toward the n-1 Gabor orientation, when the difference between this and the current Gabor orientation was 19.78^o. Again, the positive sign of a coefficient and its confidence intervals demonstrate a significant attractive bias. The width of the function is narrower than for unnormalized reports.

For trial n-2 position, there was no longer evidence for an attractive bias: the amplitude of the effect of the best-fitting derivative of Gaussian function (fitted on all current unmasked trials, with both unmasked and masked trials at n-2 position, and by using normalized response errors, zRE_n as dependent variable) was a =5668 ($-7.069*10^5$ – $7.182*10^5$), with a width of σ = 196.7^o (0 – 8741^o). The confidence intervals for a indicated that even at its peak, the effect was non-significantly different than zero.

2.5. SERIAL DEPENDENCE BY PREVIOUS GABOR VISIBILITY: ONLY UNMASKED (VISIBLE) GABORS EXERT SERIAL DEPENDENCE EFFECTS

We have established that serial dependence exists for orientation reports on unmasked, monocularly presented Gabors, in relation to the previous (n-1) presentation. The analysis, pooling all n-1 Gabor patches (unmasked and masked), obtained a serial dependence pattern similar to what has been reported in previous studies.

We then split the dataset according to the visibility of the Gabor patch presented in trial n-1, divided into unmasked (visible) and masked (invisible) past Gabors. As mentioned before, trial n was always a visible Gabor -this restriction was applied to ameliorate the noise of the analyses, for better uncovering small effects. Ambiguity regarding visibility of the past Gabors in the unmasked and masked conditions was prevented by removing data of unmasked Gabors with a PAS of 1, or masked trials with a PAS higher than 1. Note that this measure removes ambiguity only as far as self-report by the Perceptual Awareness Scale (PAS) is deemed reliable for evaluating visual awareness.

Figure 15e shows the distribution of unnormalized reports (RE_n) by $\Delta OR_{n-1,n}$, for the entire sample, split by n-1 Gabor visibility, along with the best-fitting derivative of Gaussian function and 95% prediction intervals for each condition (n-1 unmasked/n-1 masked). On visual inspection, a neat derivative of Gaussian shape characteristic of positive serial dependence effects can be observed in relation to n-1 unmasked Gabors. The coefficient values for this function are: a=11900 (3383 – 20420), σ=31.25° (20.38 – 42.12). The positive value of a and its 95% confidence interval indicates that serial dependence by n-1 unmasked Gabors is attractive and statistically significant: the peak amplitude of the effect, at 31.25° stimulus difference, was 2.95° -larger than when considering the entire sample (n-1 unmasked and masked Gabors, 1.80°). The width of the function is similar to that reported by Fischer and Whitney (Fischer & Whitney, 2014), while the size of the peak effect is intermediate between the different studies (Fischer & Whitney, 2014; Fritsche et al., 2017).

Concerning masked, reportedly unseen stimuli, the best-fitting function indicated a complete absence of attractive serial dependence, as shown by the negative value and broad confidence intervals of the amplitude a coefficient: a=-8.474*10^5 (-3.999*10^8 – 3.982*10^8), σ=376.7° (0 - 6.018*10^4). This is also shown graphically in Figure 15e, as the function (the red line) follows closely the zero line along the horizontal axis (null bias as a function of $\Delta OR_{n-1,n}$).

Similar results were obtained when reanalysing the data based on normalized repots (zRE_n), in order to remove unrelated sources of bias -as shown in Figure 15f. For n-1 unmasked Gabors, significant positive serial dependence is observed: the best-fitting DoG function (plotted in green) has a positive, statistically significant amplitude - a=217.9 (79.95 – 355.8), σ=16.42^o (11.24 – 12.61), corresponding to a peak effect of 0.1956 z-scores for an angular difference of 16.42^o with the previous Gabor.

Conversely, for n-1 masked, reportedly unseen Gabors, the best-fitting function was wider and has much lower amplitude (as seen in Figure 15e); prediction intervals show that not even at the optimal $\Delta OR_{n-1,n}$ for peak effect is the function value significantly different than zero. The coefficients of the function confirm this: a=210.9 (-162.9 – 584.6), σ=29.68 ^{o} (1.998^o – 57.36^o). In terms of z-scores of normalized response error, the amplitude of the peak effect is 0.0579, but it is not statistically significant.

For Gabors located further backwards in trial history, no significant serial dependence effect was seen, neither for unmasked nor for masked Gabors at n-2 or further positions.

In summary, there is a statistically significant attractive effect of n-1 orientation on current response (serial dependence), *only* when the Gabor patch in trial n-1 was unmasked (visible). In this condition, the tuning width and peak amplitude of the effect is similar to that seen in previous studies, particularly if the function is fitted on absolute response errors as in analyses by other authors. By contrast, when the n-1 Gabor patch was masked by continuous flash suppression (CFS), and did not reach awareness according to self-report, there is no evidence for any attractive serial dependence. The best-fitting DoG for unseen Gabors has a positive but non-significant amplitude when fitted on normalized response errors, and a non-significant negative (virtually flat) amplitude when using absolute response errors as in previous studies on serial dependence.

As a further check, we considered the case of the non-significant positive DoG that appears for n-1 masked Gabors and normalized reports (zRE_n) – see red plot in Figure 15f. It is very likely that it represents a noisy pattern without serial dependence and *not* a very weak serial dependence effect in relation to masked Gabors -especially since results for unnormalized reports, in the format utilised by other authors, are completely unambiguous on this regard (see red plot in Figure 15e). Nevertheless, we also recalled that we had obtained a weak correlation between current orientation and report for masked, reportedly unseen Gabor (as reported in section 2.2) -which would suggest an above-chance performance even when Gabors are supposedly invisible. As mentioned above, this correlation was driven by a few participants and only some specific blocks of the session. We conjectured that those participants may have actually seen some of the masked Gabors (even if they did not report it so), due to adaptation to the mask, misalignment of the dichoptic presentations, the Gabor contrast not being low enough, etc. *If* that was the case, this may have generated a very weak serial dependence pattern for supposedly unseen Gabors. Thus, we decided to apply a strict classification of n-1 Gabor 'visibility' based on objective performance. We removed all blocks with masked Gabors (and PAS=1) where there was *any* evidence for a positive correlation between orientation and response, even anecdotal evidence ($BF_{+0}>1$) – in total 7 blocks out of 120. Likewise, we also removed all blocks with unmasked Gabors (and PAS>1) with less than moderate evidence for a positive correlation ($BF_{+0}<3$) – 5 more blocks in addition to the previous 7. Furthermore, we also excluded the datasets of other two participants -12 blocks- who chose not to produce orientation reports when the Gabor was masked (and therefore performance for 'unseen' stimuli could not be assessed). Figures 15g and 15h present serial dependence analyses in relation to n-1 unmasked and masked Gabors, after excluding data on the basis of objective performance. 15g considers absolute errors, while 15h employs normalized z-scores. In both cases, as is clear on visual inspection of the plots, there is attractive serial dependence for n-1 unmasked Gabors, and complete absence of any bias for n-1 masked Gabors. This may be seen to provide further support to our conclusions. Nevertheless, we cannot objectively access Gabor visibility, and we cannot rule out that truly invisible Gabors were associated with

above-chance performance in orientation estimation. Thus, the data selection made for these analyses, based on objective performance, should be consider merely exploratory. Self-report (PAS score) remains the criterion for visibility and the key analyses are those depicted in Figures 15e – 15f and detailed in previous paragraphs.

In conclusion, our results show that serial dependence only arises if the inducing previous stimulus was processed at higher-levels of the visual hierarchy, giving rise to perceptual awareness; conversely, the effect is absent if only early sensory processing was involved. This is the case even for a low-level feature-dimension such as orientation, which is processed to a great extent at the primary visual cortex V1 (Hubel & Wiesel, 1977), and contrary to the preservation of low-level negative after-effects in similar CFS paradigms and in absence of perceptual awareness (Blake et al., 2006; He & MacLeod, 2001; Lin & He, 2009; Maruya et al., 2008). Interestingly, we did not observe a similar dissociation of regression effects by visibility (i.e. we did not observe a regression to the visible mean), as would have been expected if both effects were aspects of the same process; however, the negative findings may likely be explained by experimental constraints (particularly related to the learning process of the block statistics). At any rate, these results are consistent with our findings on serial dependence on a higher-order statistical property such as variance, where positive serial dependence was exclusively seen in presence of high-confidence decision-making about the studied dimension.

CHAPTER 5: MODELLING SERIAL DEPENDENCIES IN VISUAL VARIANCE

In our experiments on serial dependence in variance we have found evidence for two opposite history-dependent biases operating at different timescales, and likely arising at different levels of perceptual decision-making: a positive effect toward recent presentations, driven by decision processes, and a negative, longer-lasting effect, compatible with sensory adaptation. Previous studies have separately modelled these two effects: (i) attractive recency bias by iterative Bayesian operations (a process akin to a Kalman filter) and (ii) sensory adaptation by population codes subject to exposure-dependent gain changes. We built a two-layer model wherein a stimulus is encoded into a probabilistic sensory response in the lower layer (sensory layer), which is formed by population code that give rise to the negative bias. The sensory response (or likelihood probability) is then transferred to a decision layer where it is combined with a prior probability distribution containing information of all previous iterations, which is updated on a trial-by-trial basis. The model's judgment about the stimulus results from this Bayesian combination -thus explaining the positive bias to prior history. This two-layer model, based on our conclusions about the mechanisms of negative and positive perceptual biases, was able to replicate the actual serial dependences encountered in our experiments at their approximate timescales.

Our studies on serial dependence in the processing of visual variance exemplify a case of contextual modulation of perceptual decisions by previous experience. Our results as well as research works by other authors have identified two different history-dependent

biases operating at different timescales: a positive (attractive) serial dependence whereby reports about current perceptual decisions are attracted toward recent perceptual history (Bliss et al., 2017; Cicchini et al., 2014; Fischer & Whitney, 2014; Fritsche et al., 2017; Ryotal Kanai & Verstraten, 2005; Alina Liberman et al., 2014), and a negative (repulsive) after-effect driven by more prolonged exposure, or else further removed in history (adaptation after-effects) (Campbell & Maffei, 1971; Fritsche et al., 2017; Ryotal Kanai & Verstraten, 2005; Kohn, 2007; Roseboom et al., 2015).

It has been proposed that these two opposite effects have different biological roles: positive serial dependence would help to smooth away sensory noise and promote perceptual stability, whereas negative after-effects would adapt the perceptual transducer to the statistics of the current environment to maximize sensitivity around the most likely stimulus intensities (Fischer & Whitney, 2014; Fritsche et al., 2017).

Furthermore, our own findings suggest that these effects may arise at different levels of perceptual processing: apart from their different timescales, in our data the positive bias appears only in relation to high-confidence, dimension-specific decisions. Conversely, the negative bias is unrelated to decision-making and, although often becomes evident for less recent history, may even be observed for very recent presentations if the conditions for appearance of the competing positive bias are not met, such as when low confidence was reported at that point. This suggests that both effects are independent, and not an early and late stage of a single process. Classically, negative after-effects have been regarded as the product of sensory adaptation (Kohn, 2007) - although their sensory origin is in question concerning high-level dimensions (Storrs, 2015). On the other hand, our findings about serial dependence are in line with other studies proposing that it arises as decisional representations interact in working memory, and not as a bias on perception (Bliss et al., 2017; Fritsche et al., 2017; Suárez-Pinilla et al., 2018b) – but see the opposite claim in (Cicchini et al., 2017; Fischer & Whitney, 2014; John-Saaltink et al., 2016) and the possibility, raised by Kiyonaga et al. (A. Kiyonaga et

al., 2017), that serial dependence is a widespread effect arising at multiple levels of processing.

We sought to build a model capable of reproducing these two observed effects at their corresponding timescales. Based on the abovementioned insights on their likely origin, we implemented a two-layer model, formed by a sensory layer and a decision layer, that give rise to the negative and positive bias, respectively. We based the structure of the sensory layer on previous works that model sensory operations as a result of population codes subject to exposure-dependent gain changes (Heron et al., 2012; Jazayeri & Movshon, 2006; Roach, Heron, Whitaker, & McGraw, 2011). As for the decision layer, we implemented a trial-by-trial Bayesian computation with an iteratively updated prior, as has been previously employed for describing both regression effects (attraction to global history) and recency biases (attraction to recent presentations) in perceptual decision-making (Kalm & Norris, 2017; Petzschner & Glasauer, 2011).

Briefly, in each iteration, a certain **stimulus** magnitude is received on the **sensory layer** and transduced into a noisy sensory response subject to negative adaptation after-effects. This sensory response is not a single value; rather, it takes the form of a probability density function (PDF) characterizing the **likelihood** that the actual neural response had been produced by each possible stimulus magnitude.

This likelihood distribution is then forwarded to the **decision layer**. In order to translate a noisy probability function into a single perceptual judgment, decision-making is optimized by taking advantage of the information about previous perceptual history: how frequent, and therefore how likely to occur, are the different stimulus magnitudes *a priori*, i.e. before receiving any information about the current sensory input. This is also characterized by a probability function summarizing the statistics of previous iterations of the model, termed **prior** distribution. The combination of the prior and likelihood distributions in a way akin to Bayesian decision theory results in a **posterior** probability function, expressing how likely is each possible stimulus magnitude to have

actually started the current iteration, given the previous history and the current sensory response. The output of the model, i.e. its **judgment**[1] about the received stimulus magnitude, is the magnitude at which the posterior function reaches its maximum. This posterior distribution, summarizing the entire sensory history up to the current iteration, becomes the basis of the new updated prior to be combined with a new sensory response in the next iteration. Such dependence of previous history explains the decision-based positive serial dependence.

1. MODEL SUMMARY

Figure 16 presents a graphic summary of the model structure.

[1] Henceforth we will refer to subject's or model's responses as **judgments** (of the presented stimulus magnitude in each trial), to disambiguate from **population responses** of individual neural populations of the sensory layer, and from **sensory response** (which is the pooled response of all the neural populations of the sensory layer, i.e. the likelihood distribution in one trial). However, we retain the term 'response noise' for the noise corrupting the posterior distribution (and thus forming the response distribution) before judgment selection, as it is more standard than 'judgment noise'. We also employ other established terms like response preparation, response execution, etc, in relation to downstream processes responsible for managing the posterior distribution and producing a judgment.

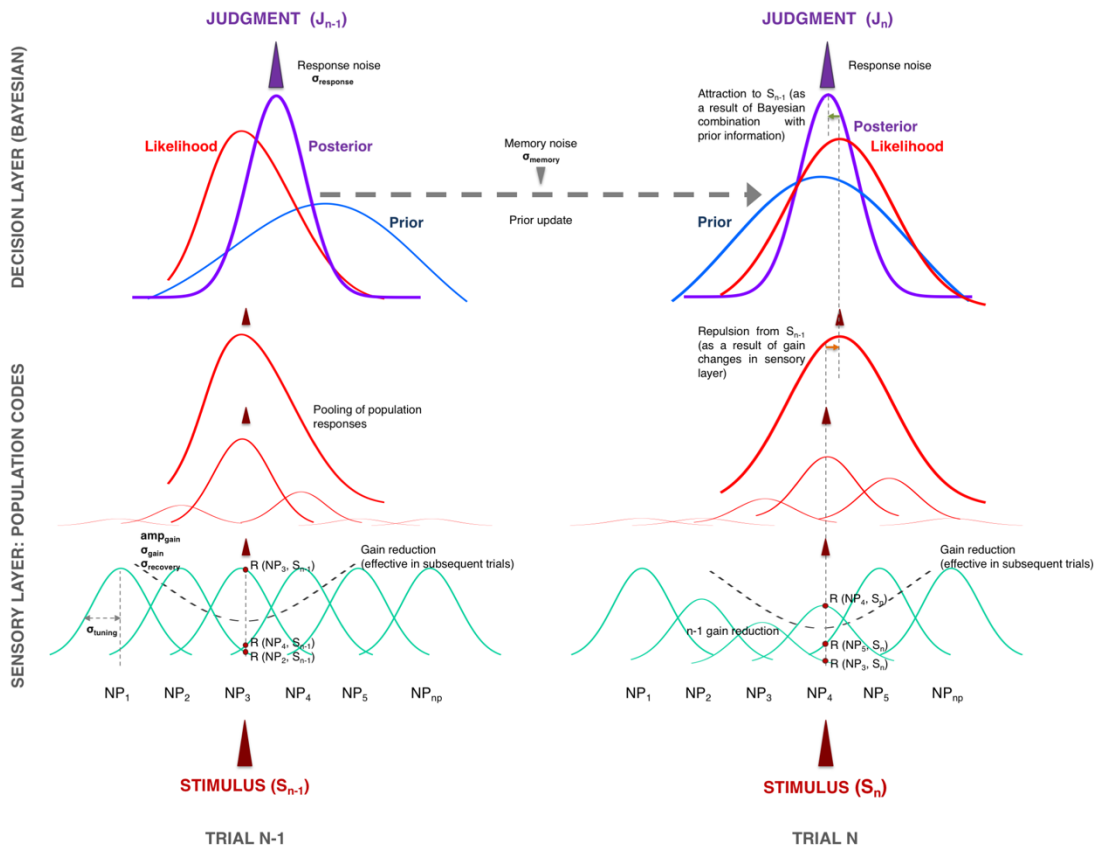


Figure 16. Model: Basic structure. The model performs perceptual decision-making across two layers: a sensory layer which processes the information from the current stimulus by neural population codes and is subject to exposure-dependent gain reduction, and a decision layer which makes an optimal judgment by a Bayesian combination of the sensory response from the lower layer and the information from the previous stimulus history, which is updated on a trial basis. Thus, negative and positive biases with respect to previous history are generated at different levels of perceptual processing and have different properties and timescales. The position of the free parameters within the model structure is detailed on the figure (see section 2.2 for further information).

1.1. SENSORY LAYER: POPULATION CODES

Negative after-effects have been successfully modelled with population codes (Heron et al., 2012; Jazayeri & Movshon, 2006; Roach et al., 2011). Population codes are a representation of the neural code (that transduces stimulus magnitudes into sensory

responses) in form of a series of **neural populations**, each sensitive to some preferred stimulus magnitudes. This preference can be modelled by associating each neural population to a **tuning function**: a probability function expressing how likely is each neuron to produce a response as a function of the stimulus magnitude it is exposed to.

Thus, the response of a neuron part of a given population (which may be expressed in terms of firing rate) will depend on the received stimulus and the corresponding value of that neural population's tuning function, but also on two other factors. The first is the internal noise of the system, which explains its probabilistic nature. The second, and key to the working of our model, is the **neural gain**, i.e. the scaling factor relating stimulus and intensity of neural response.

The actual sensory response will result of the combination of the responses from all neural populations. Because there is not an unambiguous relationship between stimulus magnitude and neural activity, this sensory response, as its components, is also probabilistic: it may be characterized by a probability function expressing the likelihood of each stimulus magnitude to have generated the actual response. In our model, this is the output of the sensory layer and is called the likelihood distribution, a Bayesian term justified by its role in the Bayesian-like computations that will take place in the decision layer.

By having neural gain subject to exposure-dependent changes, we can use population codes to model history-dependent modulation of current perception. Plus, both the selectivity of neural populations to certain stimulus magnitudes and the gain changes are biologically plausible and have been demonstrated for several perceptual dimensions (Carandini & Heeger, 2013; Dragoi, Sharma, & Sur, 2000; He, Cohen, & Hu, 1998; Kohn, 2007). Specifically, population-code models with exposure-dependent 'fatigue' (gain reduction) have been successfully employed to reproduce adaptation processes responsible for negative after-effects (Heron et al., 2012; Jazayeri & Movshon, 2006; Roach et al., 2011).

In order to model a reversal from positive to negative after-effects, a parsimonious approach might establish an exposure-dependent modulation of neural gain that caused first a short enhancement and a subsequent, more prolonged reduction. In this regard, Whitney and colleagues have modelled their serial dependence data on the basis of changes in gain or tuning of neural population codes (Fischer & Whitney, 2014). However, as stated above, our research suggests that both effects have different properties and purported origins: the positive bias likely arises from decisional rather than sensory processes, and the negative bias seems to appear almost as early as the positive effect but last for much longer, only becoming evident when the competing effect declines. For these reasons, in our model the sensory layer generates only negative after-effects, while the positive serial dependence arises from the decision layer.

1.2. DECISION LAYER: BAYESIAN-LIKE KALMAN FILTER

In our model, positive serial dependence arises from a Bayesian-like combination of the current noisy sensory response (the likelihood distribution, output of the sensory layer) and a prior probability of encountering certain stimulus magnitudes, given by trial history.

Thus, the prior distribution acts as a sort of summary memory representation, responsible for the attractive bias toward previous sensory history. As new information is received in each model iteration (i.e. each trial), this memory needs to be updated. Besides, it seems reasonable that recent sensory input should be given more weight in constructing the prior for a certain trial, due to memory limitations and the need for a balance between perceptual stability and adaptation to environmental changes. This is consistent with the attractive bias exerted by the recent trials in our data.

In order to achieve an optimal integration of information, Kalman filters provide an algorithm for recursive, Markovian update of the prior probability when both this and the likelihood are Gaussian functions, and have been used efficiently for decoding neural activity in the motor cortex (W. Wu, Gao, Bienenstock, Donoghue, & Black, 2006) or explaining other biases in perception, such as iterative estimations of displacement (Petzschner & Glasauer, 2011) and temporal regularities (Luca & Rhodes, 2016).

In the first iteration of the model, this prior is a uniform probability distribution, as the absence of previous sensory inputs, and therefore of previous knowledge about the statistics of the environment, renders all magnitudes equally likely *a priori*. In following iterations, the posterior distribution generated in the previous trial (by combination of that trial's prior and likelihood distributions) becomes the prior for the current trial after corruption by **memory noise**.

The likelihood distribution produced in the sensory layer is forwarded to the decision layer and combined with the prior in order to form the posterior distribution. The simplest case would be to assume an optimal, purely Bayesian combination, i.e. the product of both Gaussians:

(1)

$$P\left(S/R\right) = \frac{P\left(R/S\right) * P(S)}{P(R)}$$

-where R is the sensory response generated by the stimulus S; assuming all neural responses are equally likely *a priori* and therefore discarding P(R).

The resulting posterior distribution, evaluated at each stimulus magnitude within the perceptual space, represents the likelihood of each stimulus magnitude to have started the current iteration of the model, according to a Bayesian framework that takes into account previous knowledge of the environment. We assume that the model behaves

optimally, so that the selected judgment is the peak (i.e. the mean) of the posterior of the trial. However, the judgment is not selected directly from the posterior, but from the **response distribution**, which results from the posterior being transferred to areas responsible for response execution and corrupted by **response noise**. The response distribution, evaluated at each magnitude of the perceptual space, represents the likelihood of reporting that magnitude as judgment in response to the received stimulus. Response noise does not cause any bias in the response distribution, which has the same mean/peak as the posterior distribution but a larger variance -see below for detail.

As stated before, this posterior becomes the prior for next trial after corruption by memory noise.

2. DETAILED DESCRIPTION

This section is presented for completeness, as the basic structure of the model has been described above (section 1). The reader may choose to skip this part and continue in the Results section (section 3).

2.1. STIMULI

We applied the model to motion variance, a statistical feature-dimension that may be translated to a linear scale -in other words, there are lower and higher values, and two distinct ends along the perceptual space, unlike circular magnitudes like orientation or RDK motion direction. Following Petzschner et al. (Petzschner & Glasauer, 2011), stimulus magnitudes are log-transformed before entering the model:

(2)

$$S_{log} = \log(\frac{S_{lin}}{s_o})$$

where s_o=0.01 is a small normalization constant. This log-transformation represents compressive non-linear mapping of stimulus magnitudes along the perceptual scale and the power law for the relationship between stimulus magnitude and response precision.

The range of considered stimulus magnitudes (or **perceptual space**) largely corresponds to the range of allowed responses. Arbitrarily, the logarithmic perceptual space is formed by 90 evenly spaced values between the log-transformed values for 5º (log(5/0.01) = 6.2146) and 90º (log(90/0.01)= 9.1050). The model judgments can only take values that form part of the perceptual space.

The model represents information related to perceptual processing in probabilistic terms – thus, sensory response, prior information, etc, are defined as Gaussian probability density functions (PDF) computed over the perceptual space.

2.2. PARAMETERS

The model has 6 free parameters, as listed below:

1. Layer 1: sensory layer (population codes subject to gain changes that transduce a stimulus into a likelihood distribution):

    a. Parameters for gain change:

i. amp_{gain}: modulates the amplitude of the gain field of each neural population, i.e. the maximal gain reduction exerted by the activity of a neural population.
ii. σ_{gain} : modulates the breadth of the gain field of each neural population.
iii. $\sigma_{recovery}$: modulates the rate of recovery of the gain reduction, in seconds.

b. Parameters for precision of the likelihood (sensory precision):
    i. σ_{tuning} : the standard deviation of the tuning function of each neural population – i.e. the selectivity of each neural population to its preferred stimuli.

2. Layer 2: decision layer (combination of likelihood and prior distribution for Bayesian-like decision-making):

a. Parameters for precision of the prior (memory noise):
    i. σ_{memory} : memory noise added to the previous trial posterior distribution to form the current trial's prior.

b. Parameters for response precision:
    i. $\sigma_{response}$: response noise added to the current trial's posterior distribution to form the response distribution. It represents all sources of response noise occurring after the generation of a perceptual decision.

2.3. MODEL COMPUTATIONS

This section details the computations performed on a single trial (**n**), since the stimulus $\mathbf{S_n}$ is received until a judgment about its magnitude is output by the model ($\mathbf{J_n}$).

2.3.1. Sensory layer

Neural populations

The sensory layer is formed by a series of **np** neural populations NP_1, NP_2 … NP_{np}, each of them with a preferred sensitivity to a certain stimulus magnitude.

Arbitrarily, we have established that the sensory layer for the variance experiment has np = 18 neural populations, each with a maximum sensitivity ($\mathbf{S_{max1}, S_{max2}, S_{max3}, \ldots S_{maxnp}}$) to 18 stimulus magnitudes evenly distributed along the (logarithmic) perceptual space: thus, preferred stimulus magnitudes are (in log space): 6.21 – 6.38 – 6.55 … 9.11.

Tuning functions

Each neural population has a tuning curve **F** characterizing its specific (yet probabilistic) sensitivity to certain stimuli. Following Jazayeri et al. (Jazayeri & Movshon, 2006), this **tuning function** is defined per trial as the product of the **baseline tuning function**, $F_o(NP)$ by a gain factor $gain_n(NP)$ – see below for details.

Baseline tuning function

The baseline tuning function (of a neural population NP) is defined as a probability density function characterizing how likely is a neuron of that population to generate a signal in response to each stimulus of the perceptual space. Conversely, this tuning function also represents, in presence of a response from that neuron (a firing signal), the likelihood that such response has been produced by each of the possible stimuli of the perceptual space – by Bayes' rule and disregarding constant terms (the PDF is normalized so that it sums to 1).

Specifically, the baseline tuning function of NP is a Gaussian PDF defined over the whole perceptual space (see 'Stimuli' section above), with mean at the stimulus with maximum sensitivity (S_{maxnp}) and standard deviation σ_{tuning}, wherein the latter is a free parameter of the model representing sensory precision, which takes the same value for all neural populations:

(3)

$$F_o(NP) = N\left(S_{maxnp}, \sigma_{tuning}\right) = \frac{1}{\sqrt{2\pi} * \sigma_{tuning}} * e^{-\frac{\left(x - S_{maxnp}\right)^2}{2 * \sigma_{tuning}^2}}$$

for x ∈ perceptual space.

Gain

As previously stated, the tuning function of each NP is the product of this baseline tuning function by a gain factor that changes on a trial-wise basis:

(4)

$$F_n(NP) = gain_n(NP) * F_o\ (NP)$$

where the subscript n refers to the considered trial.

For the first trial of the experimental session, $gain_1(NP)=1$ for all neural populations; this value is established arbitrarily, and indicates that at the onset of the experiment we assume that the sensory layer is unbiased along the entire perceptual space.

The gain of each NP is subject to a decrease driven by the neural activity of the sensory layer, i.e., by the neural response of all NPs in response to stimuli. This gain decrease is the result of the addition of the decrease due to the activity of all (np) neural populations in response to all stimuli up to the previous trial, with each fractional effect scaled by perceptual and temporal distance.

Let us consider the gain factor of NP at trial n ($gain_n$ (NP)). It will be the sum of (n-1)*p fractional effects, where n-1 is the number of trials prior to the current one and p is the number of neural populations of the sensory layer. In each trial, a stimulus $S_{n'}$ has been received and each neural population has generated a population response R(NP', $S_{n'}$), which will have its effect on the current trial gain. Note that we are using $S_{n'}$ for the stimulus received in trial n' (n'< n) which elicits a certain response in the neural population NP'. In turn, this response will have an effect on the gain of the neural population NP at trial n and thus will modulate the response of NP to stimulus S_n, R(NP, S_n). Later on we will see how neural responses are computed. We assume that the response of the current trial does not have an instant effect on sensory gain; rather, the gain effect exerted by R(NP', $S_{n'}$) on NP is maximal after stimulus $S_{n'}$ offset and decreases afterwards. It is also maximal when NP=NP' and decreases with perceptual distance. Unlike with trials, in which we assume the current response does not instantly change the gain in the current trial, all neural populations exert an effect on NP, including NP itself –but not instantly.

Before applying temporal discount, the maximal fractional effect on $gain_n(NP)$ exerted by the neural response of NP', R(NP', $S_{n'}$), is defined as ('g_{red}' stands for 'gain reduction):

(5)

$$g_{red\_}fraction\_max(NP, NP', n') = amp_{gain} * Sdur_{n'} * R(NP', S_{n'}) * \frac{1}{\sqrt{2\pi} * \sigma_{gain}} * e^{\frac{-(Smaxnp - Smaxnp')^2}{2 * \sigma gain^2}}$$

where

1. amp_{gain} is a free parameter which scales the maximal gain reduction (before applying temporal discount, and with zero perceptual distance) and is constant for all NPs and trials.

2. $Sdur_{n'}$ is the duration of the stimulus of trial n'. Thus, we assume that the sensory response is produced in the sensory layer for the duration of the stimulus, or at least it is proportional to such duration, and the gain decrease is also proportional to the duration of the neural response that produces such effect.

3. $R(NP', S_{n'})$ is the neural response of population NP' when receiving stimulus $S_{n'}$. See below for details.

4. S_{maxnp} and $S_{maxnp'}$ are the preferred stimuli for which populations NP and NP', respectively, have maximum sensitivity; in other words, they are the stimuli at which the tuning function of the populations NP and NP' peak.

5. σ_{gain} is the standard deviation of the gain field: a free parameter that modulates the breadth of the effect of each neural response on the gain reduction of other populations.

In summary, the activity (neural response) elicited on a population NP' by the stimulus $S_{n'}$ presented in trial n' will produce a reduction on the gain of a neural population NP – we term this effect *fractional* as we are considering only the gain reduction driven by a

single neural population, NP'. This fractional gain-reduction is maximal right after offset of the stimulus $S_{n'}$ that elicited the activity, and experience time-discount afterward. The maximal fractional gain-reduction effect is positively associated with the population response triggered on NP' by $S_{n'}$ and on the duration of stimulus $S_{n'}$, and shows a Gaussian decrease with perceptual distance between the preferred stimuli of NP' and NP, scaled by the free parameters amp_{gain} and σ_{gain} (amplitude and standard deviation of the Gaussian gain function).

However, the stimulus presented in trial n' elicits a neural response not only in one neural population, but in all p populations. All of them exert an effect on NP (including NP itself). We consider that the maximal gain effect produced by the entire sensory layer on NP, as a result of the activity elicited by $S_{n'}$ is the sum of all fractional effects:

(6)

$$g_{red}\_max(NP, n') = \sum_{i=1}^{p} gred\_fraction\_max(NP, NP'_i, n')$$

This effect is maximal right after $S_{n'}$ and presents a Gaussian-shaped decrease with time. Thus, it will affect the population response to a later stimulus S_n (n>n') in an amount dependent of the elapsed time between trial n' and n. Specifically, the gain-reduction effect of $S_{n'}$ on NP at the moment in which stimulus S_n is received is in function of the maximal gain effect, the time between $S_{n'}$ offset and S_n onset and the free parameter $\sigma_{recovery}$, which modulates the decline of the gain effect with time. As the other free parameters, $\sigma_{recovery}$ takes a single value for each participant and experiment. We will explain later why this parameter is termed 'recovery'.

Formally:

(7)

$$g_{red}(NP, n, n') = g_{red}\_max(NP, n') * \frac{1}{\sqrt{2\pi} * \sigma_{recovery}} * e^{\frac{-(time(n',\ n))^2}{2*\sigma recovery^2}}$$

However, the gain of NP (in trial n) is not only affected by a single trial, but by all trials up to n. We consider that the total gain-reduction effect on NP in trial n is the sum of all gain-reduction effects due to all trials up to n (all n' $\in 1 \ldots n-1$), each one scaled by the temporal distance between that trial and n.

Formally:

(8)

$$g_{red}\_all(NP,n) = \sum_{n'=1}^{n-1} g_{red}(NP,n,n')$$

where g_{red}_all designs the gain-reduction effect exerted on a single neural population, NP, by the neural activity of all neural populations due to all the stimuli received previous to trial n. This reduction is scaled by perceptual-space distance (to the preferred stimulus of NP) and by temporal distance (of the effect of each trial in experimental history).

Because we have established before that the initial gain of each neural population is G_o=1, the gain of NP in trial n will be:

(9)

$$gain_n(NP) = 1 - g_{red}\_all(NP,n)$$

with a caveat: gain can take any value between 0 and 1 (both included), but cannot be negative. The model doesn't admit negative (inhibitory) neural responses in the sensory layer.

(10)

$$if\ gain_n(NP) < 0 \ \rightarrow \ gain_n(NP) = 0$$

Because the effect of each trial in experiment history (n' $\in 1 \ldots n-1$) declines with time, when examining the effect that one individual trial has on subsequent iterations,

we observe a recovery of the gain-reduction caused by that specific trial. This is why $\sigma_{recovery}$ is considered to measure the rate of gain recovery.

Tuning function

As a recap, the tuning function of a neural population NP in trial n is the product of its baseline tuning function by the gain factor of NP (which changes on a trial-wise basis, as we have seen):

(4)

$$F_n(NP) = gain_n(NP) * F_o(NP)$$

where $F_o(NP) = N(S_{maxnp}, \sigma_{tuning})$, a Gaussian PDF defined over the perceptual space.

Thus, the value of the tuning function of NP at one specific stimulus magnitude presented on trial n (S_n) will depend on three factors:

1. The distance between the maximum sensitivity stimulus of NP (S_{maxnp}) and the presented stimulus S_n.

2. The value of σ_{tuning}, which is a free parameter for adjusting sensory precision, i.e. internal sensory noise. This value is assumed to be equal for all neural populations.

3. The gain of that NP at trial n.

Neural response

Neural response of a single neural population (population response)

The neural response elicited on a neural population NP by a stimulus S_n presented in trial n, termed R(NP, S_n), is given by the value of the tuning function of NP at S_n:

(11)

$$R(NP, S_n) = F_n(NP, S_n)$$

As stated before, the value of the tuning function depends of the distance between S_{maxnp} and S_n, σ_{tuning} and $gain_n$(NP).

This is why σ_{tuning} is a measure of sensory precision, or internal noise: if σ_{tuning} is very small, the value of the tuning function of any neural population whose maximum sensitivity is not very close to S_n will be very low, and only the neural population that is highly tuned to that stimulus will contribute significantly to the overall response of the sensory layer: thus, the overall sensory response will be very 'crisp' and neatly tuned to the veridical stimulus. Conversely, if σ_{tuning} is large, the value of the tuning function of most neural populations will be large enough to contribute significantly to the overall response, even if their peak sensitivity is far from the current stimulus level. In this case, the overall sensory signal will be less precise – less informative regarding the stimulus that originated it.

For simplicity, our model does not consider external noise: each RDK provides one unambiguous variance stimulus, given by the dispersion of the direction of its components –even if variance itself is a statistical property related to the precision of an ensemble.

Overall sensory response – likelihood distribution

Following Jazayeri et al. (Jazayeri & Movshon, 2006), the logarithm of the probability distribution representing the neural response elicited by the entire sensory layer (when receiving S_n) is given by the weighted average of the logarithms of the tuning functions of all neural populations, each one weighted by its neural response.

Formally:

(12)

$$\log\big(R(S_n)\big) = \sum_{i=1}^{p} R(NP_i, S_n) * \log\big(F_n(NP_i)\big)$$

This probability distribution is a function defined along each considered value of the perceptual space. We now normalize it by the sum of all its values along the perceptual space so that it sums to one:

Let ps be the number of considered values in the perceptual space (90 in our experiment), x_i each of those values and SumR(S_n) the sum of all values of R(S_n) for all x_i in the perceptual space:

(13)

$$\mathrm{SumR(Sn)} = \sum_{i=1}^{ps} R(S_n)(x_i)$$

However, previously we have calculated log(R(S_n)) instead of R(S_n) directly.

(14)

$$\mathrm{SumR}(S_n) = \sum_{i=1}^{ps} R(S_n)(x_i) = \sum_{i=1}^{ps} e^{\log\,(R(S_n)(xi))}$$

Therefore:

(15)

$$\log\left(R_{norm}(S_n)\right) = \log\left(\frac{R(S_n)}{SumR(S_n)}\right) = \log(R(S_n)) - \log\left(SumR(S_n)\right) = \log\left(R(S_n)\right) - \log\left(\sum_{i=1}^{ps} e^{\log\left(R(S_n)(xi)\right)}\right)$$

The normalized response of the entire sensory layer after reception of a stimulus S_n is what we call the likelihood distribution: $Lk_n = R_{norm}(S_n)$, representing the probabilistic sensory response to stimulus S_n, defined at each value of the perceptual space.

(16)

$$Lk_n = R_{norm}(S_n) = e^{\log\left(R_{norm}(S_n)\right)}$$

The mean and standard deviation of the likelihood distribution are calculated weighting values by their respective probability:

(17)

$$\mu_{Lkn} = \sum_{i=1}^{ps} Lk_n(x_i) * x_i$$

(18)

$$\sigma_{Lkn} = \sqrt{\sum_{i=1}^{ps} Lk_n(x_i) * \quad (x_i - \mu_{Lkn})^2}$$

In normal conditions (i.e. unless gain-related parameters take highly deviant values), the likelihood distribution can be approximated to a Gaussian shape:

(19)

$$Lk_n \approx N(\mu_{Lkn}, \quad \sigma_{Lkn})$$

For simplicity, in the next steps we will treat the likelihood as a Gaussian probability density function computed over the perceptual space.

This likelihood distribution is the neural response generated in the sensory layer by the stimulus S_n: in other words, the output of the sensory layer as produced by the model, which is then transferred forward to the decision layer.

2.3.2. Decision layer

Computation of the posterior distribution

The stimulus presented in a given trial n (S_n) is not directly accessible by the visual system: only the noisy sensory response that such stimulus has produced in the sensory layer is accessible –i.e. the likelihood probability distribution. As previously stated, this probability represents how likely is the actual neural response to have been produced by each possible stimulus magnitude of the perceptual space.

Within a Bayesian framework, the optimal decision about the stimulus magnitude must combine the noisy sensory information of the current trial with the knowledge of which stimuli were presented in previous iterations. The latter is conveyed by the prior distribution, which represents the probability of encountering different stimuli given the experimental history. This prior distribution is stored in the decision layer and continuously updated with the information provided by each new trial.

The posterior distribution is the probability function resulting of the combination of the likelihood and prior distributions. It represents the probability of each value of the perceptual space to have been the presented stimulus magnitude, given the sensory information *and* the previous history.

At the beginning of the experiment (trial n=1), in absence of previous information, all stimuli are deemed equally likely *a priori*; therefore, trial 1 has a flat (uniform) prior defined over the perceptual space. Since the relative probabilities of each perceptual

magnitude are unaffected by the flat prior, the posterior of the first trial is equal to the likelihood:

(20)

$$Post_1 \approx N(\mu_{Lk1}, \quad \sigma_{Lk1})$$

All probabilities are normalized so that their sum is always 1, so the combination with a flat prior does not change the value of the function at each value of the perceptual space.

In subsequent trials, however, the prior can be approximated to a Gaussian probability density function computed over the perceptual space:

(21)

$$\text{Prior} \approx \text{N}(\mu_{Prior}, \quad \sigma_{Prior})$$

We will later see how the prior is computed.

The posterior distribution is obtained from an optimal Bayesian combination of likelihood and prior. The result of this computation is a Gaussian probability density function whose mean is a weighted average of the mean of the likelihood and prior (Petzschner & Glasauer, 2011):

(22)

$$\mu_{Post} = W_{Prior} * \mu_{Prior} + (1 - W_{Prior}) * \mu_{Lk}$$

where

(23)

$$W_{Prior} = \frac{\sigma_{Lk}^2}{\sigma_{Prior}^2 + \sigma_{Lk}^2}$$

and

(24)

$$W_{Lk} = 1 - W_{Prior} = \frac{\sigma_{Prior}^2}{\sigma_{Prior}^2 + \sigma_{Lk}^2}$$

Thus, the mean of the posterior distribution is a weighted average of the mean of prior and likelihood, wherein each one is weighted by the variance of the other distribution. This means two things:

1. The bias exerted by previous history at the decision layer is always attractive, since the mean of the posterior will be an intermediate value between μ_{Lk} and μ_{Prior}. In our model, repulsive effects are generated at an earlier stage of processing, by exposure-dependent gain-decrease in the sensory layer.

2. Reliance on previous history depends on the precision of the current sensory signal (a more precise sensory signal implies lower σ_{Lk}^2 and W_{Prior}) and of the previous history (which might be considered memory precision: the more precise, the lower σ_{Prior}^2 and the larger W_{Prior}). In other words, reliance on previous history will be stronger when the current sensory signal is highly imprecise, or when the memory representation of previous history is highly precise.

The variance of the posterior is also dependent on the variance of prior and likelihood:

(25)

$$\sigma_{Post}^2 = \frac{\sigma_{Prior}^2 * \sigma_{Lk}^2}{\sigma_{Prior}^2 + \sigma_{Lk}^2}$$

The posterior distribution has two roles in the model: providing the basis for response selection and the basis for the new, most updated prior for the next trial. In order to fulfill these two roles, we assume the posterior is, on the one hand, transferred to downstream areas responsible for response preparation and execution, and on the

other hand, stored within the 'memory' of the decision layer. However, the posterior does not come out of these processes intact: it is corrupted by two separate sources of Gaussian noise: response noise and memory noise, respectively.

Judgment

The posterior distribution provides the basis for response selection. However, the response (judgment) is not extracted directly from the posterior distribution. As stated in the previous section, the posterior is transferred to downstream areas responsible for response preparation and execution, where it suffers corruption by Gaussian response noise with properties $N(0, \sigma_R)$. This 'corruption' is mathematically expressed by convolution of two Gaussians: the posterior distribution and the Gaussian response noise. The resulting distribution (henceforth called the response distribution) is a Gaussian probability density function with properties:

(26)

$$Resp_n = N(\mu_{Postn}, \sqrt{\sigma^2_{Postn} + \sigma^2_{response}})$$

The measure of response noise $\sigma_{response}$ is a free parameter in our model.

The model's judgment about stimulus S_n is obtained from this response distribution: specifically, it corresponds to $\mu_{response} = \mu_{Postn}$, i.e. to the value were the response distribution peaks. Thus, our model always selects the optimal response, the most likely value to correspond to the presented stimulus, after combining the current sensory information and the previous history. Note that the response noise does not produce any bias, since the peak/mean of the response distribution is the same as the peak of the posterior. However, it increases the width of the distribution, with respect to the posterior, and therefore, reduces the difference between the most likely and less likely values, so that non-optimal responses (i.e. different from the peak of the distribution,

such as those that are performed by real subjects) will be more likely under the present conditions than if the response was based directly on the posterior. Thus, response noise does not affect model judgments, but it is relevant for parameter selection through maximum likelihood estimation (see below).

Prior update

As stated before, the posterior distribution is also stored in the decision layer, where it will become the basis of the prior for the next trial, representing the most updated knowledge of the environment given by the summary of all previous stimuli, with greater weight for the most recent ones. However, between the current trial and the next one, this posterior is corrupted by Gaussian memory noise, with the shape $N(0, \sigma_{memory})$, so that the prior of the next trial will be a Gaussian distribution resulting of convolution of posterior and memory noise:

(27)

$$Prior_{n+1} = N\left(\mu_{Postn}, \sqrt{\sigma_{Postn}^2 + \sigma_{memory}^2}\right)$$

σ_{memory} is the sixth free parameter in our model. Similar to the case of response noise, the prior is unbiased with respect to the posterior of the previous trial ($\mu_{Priorn+1} = \mu_{Postn}$), but because of memory corruption it is less sharp than the posterior.

2.4. PARAMETER SELECTION

The values for the six free parameters (amp_{gain}, σ_{gain}, $\sigma_{recovery}$, σ_{tuning}, σ_{memory}, $\sigma_{response}$) are selected based on which combination of values has maximum likelihood given the model and the actual data.

As previously indicated, the likelihood of a judgment J_n produced by a participant in a given trial, given the model and a specific set of parameter values, corresponds to the value that the response probability density function (defined along the perceptual space) takes at J_n. In turn, the likelihood of the whole sequence of judgments provided in the experiment (i.e. the likelihood of the entire dataset) will be the product of the likelihoods of the individual trials:

(28)

$$P(data/parameters) = P(J_1 \cap J_2 \ldots \cap J_n) = P(J_1) * P(J_2/J_1) \ldots * P(J_n/J_1 \cap J_2 \ldots \cap J_{n-1}) = Resp_1(J_1) * Resp_2(J_2) \ldots * Resp_n(J_n)$$

where J_n represents the participant's judgment in trial n, $Resp_n$ is the response PDF produced by the model for trial n, and $Resp_n(J_n)$ is the value of that PDF at J_n. We can treat the conditioned probability as a product of these functions' values, since the response function for each trial, as defined before, is conditioned to all previous trials (represented by that trial's prior).

By Bayes'rule we can obtain the likelihood of a specific set of parameter values:

(29)

$$P(parameters/data) = P(data/parameters) * P(parameters)/P(data)$$

Since all possible parameter values are considered equally likely *a priori*, and the data (the actual responses) is the same for all tested sets of parameters values, we can dismiss P(parameters) and P(data) and conclude that P(data/parameters) will be maximal for the same values than P(parameters/data).

The parameter values that are tested in each case are:

1. amp_{gain}: [0, ∞), the sensory layer is only allowed for negative (repulsive) effects (amp_{gain}>0, or else for absence of effect amp_{gain}=0).

2. σ_{gain}, σ_{tuning}, $\sigma_{response}$: from 0 to the higher end of the perceptual space (90º).

3. σ_{memory}: from 0 to infinite, since we allow the possibility of complete lack of memory, equivalent to a flat prior.

4. $\sigma_{recovery}$: from 0 to 60 (seconds).

For each considered dataset, the parameters with maximum likelihood are determined by using the Matlab function fminsearchbnd.

Before parameter selection, our experimental data was rescaled to remove systematic biases. The rationale of this provision was as follows. For translating the position of the response bar along a visual analogue scale into variance reports, in our data analysis we employed a linear translation, so that both ends of the response scale corresponded to the lowest and highest RDK standard deviation presented during the training (0º and 90º), and the selected position was assigned a numeric value given the linear distance between both ends (for example, the middle point corresponded to 45º). These values pertained to an unfamiliar dimension, namely the standard deviation of the von Mises distribution for the direction of the individual dots in the RDK. Given the abstract and unfamiliar nature of the judged dimension, the use of a visual analogue scale and the conventional linear translation, it is not surprising that participants' average judgments, once translated into a number, deviated from the veridical standard deviation, with a marked trend to 'overestimate' large StD values - especially since the maximum StD employed in experimental blocks was 60º, leaving one third of the response scale toward the right end free for such apparent overestimation (see Figure 2a). While this conventional linear translation was valid for analysing trial-by-trial biases in participants' normalized judgments, our model was not designed to account for systematic biases. As a result, using the unmodified reports would have rendered many judgments highly unlikely under most combinations of parameter values -since most combinations would

produce judgments centred around the veridical value. For this reason, we removed these apparent systematic biases (given by our numerical translation of participant's judgments) by subtracting the difference between each participant's mean response for each StD and the veridical StD value. Parameter selection was based on these re-centred judgments.

3. RESULTS

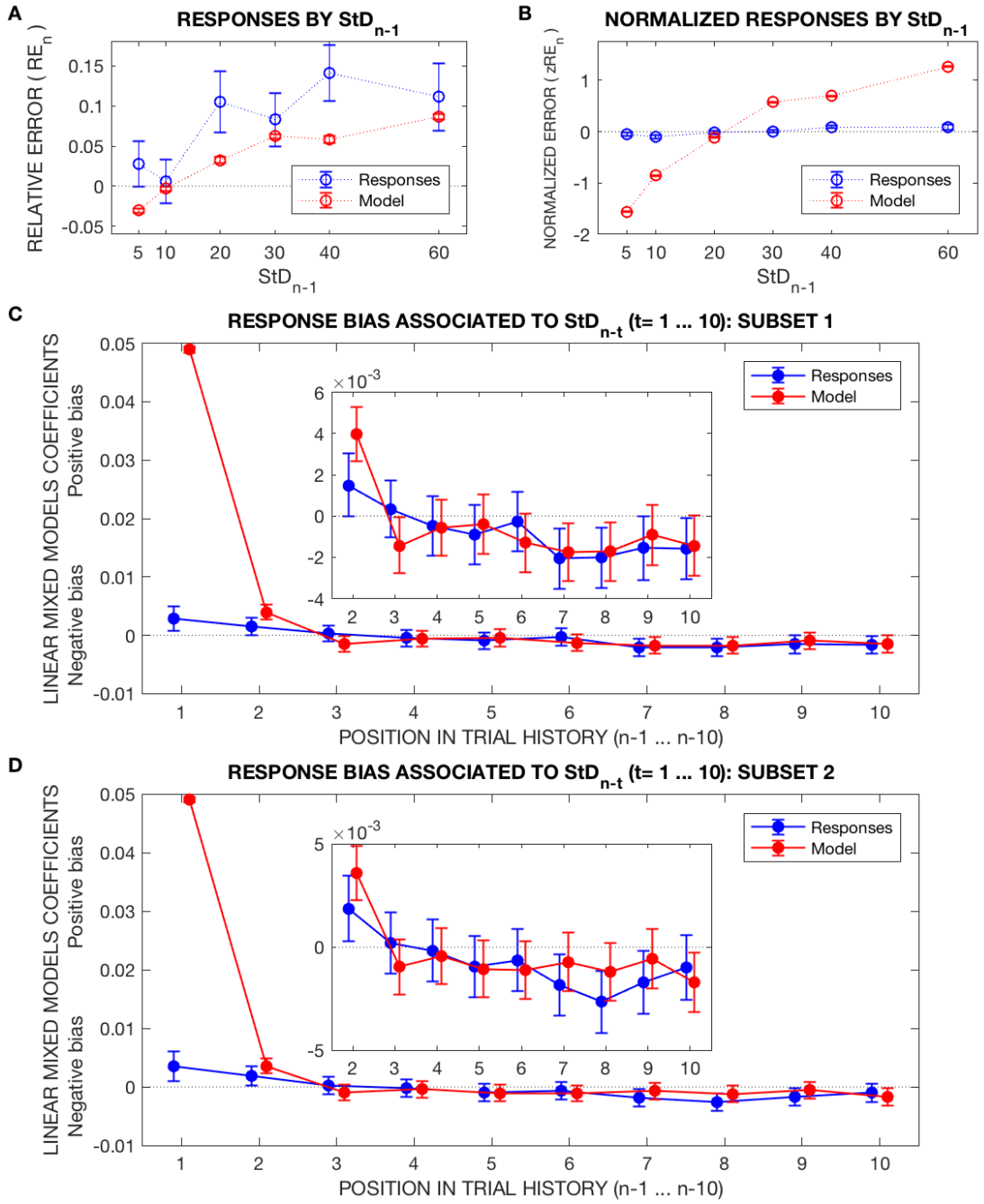


Figure 17. Model: Results. A pattern of recent positive and less recent negative serial dependencies is observed for the model outputs as well as for participants' responses. **17a - c** present data corresponding to half the participants that took part in Experiment 1 (subset 1, formed by 15 participants, randomly selected out of 30). Model judgments

are based on parameter values fitted for the data of subset 1. **17a.** Average relative errors made in the current trial (n) as a function of the variance presented in trial n-1 (StD_{n-1}). Relative errors are defined as $RE_n=(R_n-StD_n)/StD_n$, where R_n is the reported variance judgment. The blue plot represents participant's, while the red plot represents model judgments. In both cases, error bars indicate the between-participant standard error. An ascending slope is observed for both plots, indicative of an attractive bias related to StD_{n-1}. **17b.** Average relative errors, normalized within participant and current stimulus (StD_n), as a function of StD_{n-1}. Again, the ascending slope of both plots (responses, outputs) indicates positive serial dependence, although, due to the smaller variability in model outputs compared to real responses, the magnitude of the bias appears much larger for the former. **17c.** Fixed-effects coefficients for Bayesian LMMs for the association between StD_{n-t} (t = 1 … 10) and normalized response or model output in the current trial. Positive and negative serial dependencies are seen at similar timescales for participants' and model judgments. The inset shows detail for the positions n-2 to n-10. **17d.** Analogous analyses as presented in **17c**, but applied to the other 15 participants of Experiment 1 dataset (subset 2). Model judgments are obtained by using the parameter values fitted for subset 1.

For parameter fitting, we randomly selected a subset of half (N=15) the participants that took part in Experiment 1 (serial dependence in variance judgments, reported in Chapter 1). In the manner described in section 2.4 of the current chapter, 'Parameter selection', we obtained the parameter values that led to the maximum likelihood given the actual data of this subset (henceforth termed subset 1) and the structure of the model. We then ran the model with the fitted parameters on the same trial sequences (the experimental session of the 15 participants in subset 1) and obtained the corresponding model judgments. Finally, we analysed both the participants' and the model judgments for serial dependencies driven by the stimuli presented in past trials, up to n-10. The methodology of these analyses is the same that we followed for experimental data (Chapter 1).

Figures 17a-c summarize serial dependence analyses on subset 1. The blue plots correspond to participants' actual judgments (after subtraction of systematic, average bias, see section 2.4), and the red plots to the model's predicted judgments. Figure 17a presents the average relative response error (defined as $RE_n=(R_n-StD_n)/StD_n$) as a function of the StD presented in the previous (n-1) trial, StD_{n-1}. Both plots, corresponding to participants and model judgments, exhibit an ascending slope, indicating an attractive effect on current judgment in relation with the previous stimulus. In Figure 17a

participants and model judgments and unnormalized; on visual inspection, the size of the attractive bias (represented by the slope of the plot) seems similar for actual responses and model outputs. By contrast, when judgments are normalized within participant and current StD_n, as presented in Figure 17b, the effect size appears to be much larger for the model. This apparent discrepancy is explained by the much lower variability in model judgments compared to participants' real data - since the model always selects the optimal judgment, i.e. the peak of the posterior probability distribution. Thus, z-scores in normalized model outputs represent a smaller increase in terms of unnormalized values, compared to participants' responses.

In a similar manner than for experimental data (see Figure 2d for example), we analysed serial dependence on model outputs by Bayesian LMMs with normalized judgment as dependent variable, and StD in a past trial position as independent variable. Figure 17c presents the fixed-effects coefficient estimates of 20 (10x2) Bayesian LMMs for participants' normalized response (zRE_n) -blue plot- *or* normalized model judgment -red plot- as dependent variable, evaluating the effect of StD_{n-t} (t = 1...10, each position assessed in a separate model), with random effects grouped by participant's ID. The coefficient estimates represent the increase (in z-scores) observed in normalized responses or model judgments for each 1^o of increase in previous (n-t) StD. A positive bias driven by previous StD is observed in relation to positions n-1 and n-2, while a negative effect arises for more remote positions, mainly n-7 and n-8. The timescale of both types of history-dependent bias is similar for real participants' and model judgments. Although the positive effect driven by n-1 trial appears much larger for model outputs (B= 0.0489, 95% credible intervals (0.0484 - 0.0494)) compared to participants' data (B= 0.0029 (0.0008-0.0049)), this is explained to a great extent by the differences in normalized scores, as shown in Figures 17a and 17b. On the contrary, the negative effect by less recent history seems of a similar magnitude for normalized responses and model outputs, but is weaker for the latter if unnormalized outputs are considered. This negative effect is statistically significant at positions n-3, n-7 and n-8, peaking for StD_{n-7}: B=-0.0018 (-0.0032 - -0.0003). The inset graph in Figure 17c presents

the same results as the main plot, leaving aside position n-1 to better appreciate the effect size of serial dependence from n-2 to n-10. The progression of the positive and negative serial dependencies (considering normalized reports) is similar for real data and model outputs.

We then enquired whether the same parameter values (fitted for subset 1 data) could produce similar serial dependencies on a different trial sequence. Thus, we ran the model on the remaining 15 participants of Experiment 1 dataset (subset 2), and analysed serial dependence in relation to the obtained model judgments. This analysis is depicted in Figure 17d, analogous to Figure 17c but concerning subset 2 data. A similar pattern of positive and negative serial dependencies is observed in this case.

In conclusion, we built a model on the basis of our conclusions about the origin of the two opposite history-dependent biases in perceptual judgments that were observed in our experimental data: negative biases of likely sensory origin and positive serial dependencies of presumed decisional basis. Operationalized in a similar manner as previous models of sensory adaptation and Bayesian decision theory, respectively, these biases are obtained for the model outputs in a similar magnitude and timescale as encountered for human participants.

CHAPTER 6: DISCUSSION

In this chapter we discuss the research findings reported in previous chapters of Part II of this thesis. Our experiments on serial dependence in visual variance have revealed the existence of two different history-dependent biases that may be observed in the same experimental paradigm: a positive bias in relation to very recent presentations, and a negative effect toward a broader, less recent context. Results indicate that the positive serial dependence is driven by high-confidence dimension-specific decisions (and not by perception per se*) and likely subject to time and capacity limitations that may suggest the involvement of memory processes. A continuous flash suppression experiment about serial dependence in orientation further demonstrates that, even for low-level feature-dimensions, serial dependence cannot arise from local changes in sensory areas -contrary to negative adaptation. Finally, a model based on our conclusions about the basis of the two opposite history-dependent biases replicated the observed serial dependencies in their approximate timescales. In summary, we argue that perceptual decision-making about visual variability (and likely many other visual dimensions) is subject to a combination of negative sensory and positive decisional biases in order to optimally tune the visual system to balance the need for change sensitivity and perceptual continuity in a largely stable environment. This chapter draws extensively from the discussion in the Journal of Vision article 'Serial Dependence in the Perception of Visual Variance' (Suárez-Pinilla et al., 2018b).*

The examination of serial dependence provides a valuable window on perceptual processing. In a series of studies, we applied this approach to visual statistics rather than

to individual perceptual features: specifically, to variance, a basic trait in the interpretation of noisy information about complex visual scenes. We found evidence for two opposite serial dependence effects operating on different timescales: an attractive (positive) bias associated with very recent variance presentations, which is exerted only when a judgment about that dimension was made in the most recent 1-2 trials and high confidence was placed in that decision, and a repulsive (negative) bias which appears even for the most recent trial history for low-confidence variance presentations, but which generally becomes manifest several trials into history and persists for at least ten trials.

Several studies on serial dependence have found a positive (attractive) bias towards recent perceptual history, which is modulated by attention, enhanced by spatial proximity yet not specific to retinal location, takes place in the fovea as well as the periphery, and fades after 5-15 seconds but does not require explicit memory (Fischer & Whitney, 2014). While control experiments support that this effect does not require a motor response, there is an ongoing debate about whether its basis is perceptual or post-perceptual: the results of a two-alternative forced-choice discrimination task (Fischer & Whitney, 2014) (with a sample size of three participants), a recent behavioural study (Cicchini et al., 2017) and a V1-based fMRI study (John-Saaltink et al., 2016) have been used in support of a perceptual origin, while another study employing a combination of appearance and performance tasks has made the case for a post-perceptual (decisional) source (Fritsche et al., 2017). All four studies examining the mechanistic basis of serial dependence have used a low-level feature like orientation; nevertheless, serial dependence has also been described for high-level features, including facial appearance (Alina Liberman et al., 2014; Xia et al., 2015), relative timing (Roseboom, 2017) and statistical properties such as numerosity (Cicchini et al., 2014) and ensemble mean (Manassi et al., 2017).

In our experiments on visual variance (a high-order visual statistic), we found a positive bias that shares many of the characteristics listed above, but differs in others. In terms

of similarities, it operates on a similar timescale (temporal tuning seems to be slightly shorter for high-level domains, as shown in a study with face perception (Alina Liberman et al., 2014) and also in our data), occurs similarly across presentation eccentricities, and is not related to response execution. In addition, it seems to be non-linearly tuned by the (di)similarity between previous and current stimulus, although results may be confounded by the particularities of our task. On the other hand, it exhibits other characteristics that suggest that, for visual variance, the bias depends on decisional rather than perceptual processes. First, it is entirely independent of retinal location, appearing with similar magnitude for successive stimuli displayed at the same position or at an angular distance of 40^{o} – as shown in the peripheral trials in Experiment 1. Second, it is independent of a closely related statistical property - the mean direction (previous studies have highlighted a strong relationship between mean and variance, showing that variance plays an important role in the accuracy and confidence of mean judgments) (Fouriezos et al., 2008; John Maule & Franklin, 2015). Together, these properties make a low-level, perceptual origin very unlikely. Note that priming of mean judgments by visual variance, as described in (Michael et al., 2014), is also independent of the similarity of means and retinal location.

The most compelling argument in favour of a decisional origin for the positive serial dependence in our results is that, in a task-switching design, the bias disappears entirely when, for the past stimulus (i.e. the inducer), participants were engaged in a decision about a different feature-dimension than variance. This is repeatedly shown in Experiments 2B-2E, where participants make decisions about either the variance or the mean direction of the RDK stimuli. It is particularly notable, since mean judgments are strongly dependent on ensemble variance (Fouriezos et al., 2008; John Maule & Franklin, 2015), and the stimulus is identical for both tasks. Furthermore, this dissociation persists even when the required task was post-cued (as in Experiments 2D and 2E), ruling out an alternative explanation involving differences in feature-specific attention. According to previous studies, serial dependence is enhanced by attention (Fischer & Whitney, 2014), and a pre-cued task-switching experiment may have caused

differences in perceptual attention to mean and variance in function of the cued task. However, when the cue only shows after stimulus removal, any difference involving both tasks must be of post-perceptual origin. In fact, in our entire series of task-switching experiments, only in one instance there was serial dependence in variance in relation to past trials where a decision about mean was cued: this occurred when a long interval (2.5 seconds) elapsed between stimulus offset and task cue, and presumably participants had time to prepare both possible decisions (mean and variance) although only a mean report was eventually required. In summary, results of our task-switching experiments show that serial dependence is generated by past, dimension-specific perceptual decisions.

Experiment 3 places a further constraint on what is transferred in serial dependence: not *any* dimension-specific decision, but specifically those that attain a certain level of confidence. From a predictive perception perspective, the fact that only high-confidence trials drive the positive serial dependence may be considered supportive of both perceptual and decisional origin, as a more precise prior would give rise to a stronger reliance on sensory/decisional history (Meyniel et al., 2015). However, an interpretation based on sensory precision might also predict two associations that are not found in our data: (i) an inverse association of positive serial dependence with current-trial confidence, and (ii) an inverse association with eccentricity, given lower sensory precision in the peripheral field. Rather, our experiments strongly support a lack of association of serial dependence with these two factors. In broader terms, serial dependence in variance judgments could be regarded as part of a generic strategy of mirroring or transferring trusted decisions -indeed a recent study has found that confidence boosts serial dependence even when dissociated from task performance (Samaha, Switzky, & Postle, 2018). This explanation could also encompass the negative serial dependence associated with low confidence (as a repulsion away from judgments deemed unreliable); however, the different timescales over which the positive and negative biases operate suggest that they are independent mechanisms rather than two aspects of a confidence-based strategy (Alais, Ho, & Han, 2017).

In summary, it is likely that positive serial dependence in variance is driven by high-level decision-making processes. In this respect, our findings agree with Fritsche et al. (Fritsche et al., 2017), who assert the same for orientation judgments. These authors propose that working memory representations are biased toward previous (dimension-specific and task-specific) decisions, a hypothesis that is supported by the potentiation of the bias when several seconds are allowed between stimulus offset and response. Bliss and colleagues (Bliss et al., 2017) provide converging evidence, reporting that serial dependence is absent at the moment of perception but increases in visual working memory, reaching a maximum when a 6-second delay between stimulus offset and response is placed (a similar study, however, has reported evidence for serial dependence at the time of perception (Manassi et al., 2018)). Interestingly, in 2005 Kanai and colleagues also found a positive bias on the reported direction of ambiguous motion, appearing in relation to a past reported percept (and not to the low-level sensory signal) only when the stimulus was presented several seconds after the adaptor; they called this effect 'perceptual sensitization' (Ryotal Kanai & Verstraten, 2005).

In our data we also found evidence suggesting that serial dependence is subject to time effects and capacity limitations, superficially akin to those observed for working memory representations (G. A. Alvarez & Cavanagh, 2004; T. F. Brady, Konkle, & Alvarez, 2011; Huang & Sekuler, 2014). Regarding potential time effects, similar to reported findings by Bliss (Bliss et al., 2017) and Fritsche (Fritsche et al., 2017), we observed a trend toward *increased* serial dependence effects when a longer time interval is interposed between stimulus and response, possibly in relation to a greater influence of past representations on *current* decisional information as this is kept for longer time in working memory. In addition, we also found a trend toward *weaker* serial dependence for longer inter-trial intervals (time during which participants are passively waiting for the next trial), which might suggest a wearing-down effect of a *past* decisional representation stored in memory. These two distinct effects of time on the magnitude of serial dependence are consistent with a description of memory-driven biases in terms

of Bayesian operations (Kalm & Norris, 2017; Olkkonen et al., 2014; Raviv et al., 2012), since, due to memory noise, a longer interval before response would render the current information (likelihood) less precise, whereas a longer time between trials would lower the precision of the prior, tuning the relative weight of prior information (serial dependence) in the directions that were actually observed in each case.

We also encountered evidence suggesting that serial dependence is affected by capacity limitations, again showing similarities with mnemonic processes. Specifically, we found that positive serial dependence is disrupted by additional decision-making along the same feature-dimension. In other words, the effect of remote (n-2) variance decisions on current (n) judgments is disrupted by interposing (n-1) decisions about variance, whereas it is noticeably larger if the n-1 decision was about a different feature-dimension (e.g. mean). Such difference appears while keeping time constant and for both pre-cued and post-cued tasks. Albeit not statistically significant, this result suggests the existence of a dimension-specific capacity-bottleneck, where decisional representations about a certain feature-dimension merge and overwrite previous, similar representations (Matthey et al., 2015). The different effects of both decision types (which arguably demand a similar amount of resources) in function of the similarity with other decisions would be easy to interpret in light of a resource-based model of working memory, where the latter is not represented by a fixed number of independent slots, but as a fluid resource that can be flexibly allocated to construct hierarchical memory representations over time (Suchow et al., 2014). Furthermore, we uncovered converging evidence (yet also non-significant) in post-hoc analysis of the differences in the pattern of serial dependencies in Experiments 1 and 3. These experiments were identical in design except for an additional requirement of a confidence report in the latter, which also produced longer times between consecutive stimuli. Both factors (time and the additional decision) showed a negative interaction with serial dependence - even though the degree of similarity between a decision about variance and an additional decision about confidence in variance estimation is open to debate (Adler & Ma, 2016; Kepecs et al., 2008; Meyniel et al., 2015; D. Rahnev et al.,

2015). Nevertheless, since our results on time and capacity constraints were not statistically significant, the factors contributing to the disruption or fading of positive serial dependence in relation with more remote presentations are deserving of further research.

All in all, our results suggest that the mechanisms responsible for the generation of serial dependence would also be the cause of its fading: both would be two aspects of a merging process due to interaction of successive memory representations about perceptual decisions, driven by time and capacity limitations. While the specific mnemonic processes involved are unclear - our methodology was not designed to operationalize specific instances of memory (such as working memory) -, in line with our observations, Kiyonaga and colleagues have noted the similarities between serial dependence effects and well-studied disruptions related to working memory limitations (such as proactive interference) (Anastasia Kiyonaga, Jason M. Scimeca, Daniel P. Bliss, & David Whitney, 2017); these authors suggest that the latter might be a maladaptive aspect of a generally beneficial and widespread brain mechanism for stabilizing internal representations at different levels of processing, including perception, attention and memory.

While our results clearly demonstrate that a past decisional representation (and not merely attentive perception) is necessary to induce serial dependence in subsequent variance judgments, one could argue that still the *induced* bias could be partly exerted at a perceptual level. In this respect, as also pointed by Kiyonaga et al. (A. Kiyonaga et al., 2017), working memory encodes information by attentional recruitment of internal representations stored in the same areas responsible for non-mnemonic processing (D'Esposito & Postle, 2015): for instance, sensory areas are engaged by visual working memory tasks (Baumann et al., 2008). Thus, it might be hypothesized that serial dependence represents an effect of working memory representations on the activity in visual areas, capable of altering subsequent perception. In line with this, the finding of fMRI activity signal in the primary visual cortex (V1) in relation to the previous trial

choice (*not* the previous stimulus) in an orientation task (John-Saaltink et al., 2016) may be supportive of serial dependence as a top-down bias by previous history which feeds back to early sensory areas. On the other hand, Fritsche and colleagues found that tasks depending strictly on perception (scarcely subject to decisional or mnemonic biases), like a perceptual comparison (2AFC) or an equality task, were not subject to serial dependence by previous decisions, demonstrating, according to these authors, that serial dependence does not affect perception *per se* (Fritsche et al., 2017) (but see opposite claim in (Cicchini et al., 2017)). Fritsche et al. cite supporting evidence by a fMRI study showing that perceptual hysteresis (stabilization toward previous perceptual reports) maps into higher-order visual and fronto-pariental areas, contrary to adaptation processes involving early visual areas (Schwiedrzik et al., 2014). In our own study, we did not deem suitable to employ a perceptual comparison or equality task to ascertain the existence of perceptual biases in variance, since serial dependence effects were independent on location and other low-level stimulus properties: thus, any potential bias on current perception would have equally affected both items to be compared. Nevertheless, the fact that serial dependence in our experiment is driven by a past decision (never only by perception or memory of perception) points toward a direct interaction between past and current decisional representations, likely recruited by working memory in high-level decision-related areas, without any explanatory power left for induced changes in perception.

It may be proposed that our results, indicating that serial dependence is driven by past decisions and not low-level perception, may be restricted to abstract, high-order properties such as visual statistics. Indeed, our findings are suggestive of a high-level mechanism of variance processing, not only concerning serial dependence (as discussed extensively above), but also other aspects of variance estimation. For example, in Supplementary Experiment 1B (see Chapter 1, section 3.3), performance in variance estimation (measured by response times, accuracy or response dispersion) was independent of ensemble mean. A putative high-level locus for variance is also in agreement with the conclusions of Payzan-Le Nestour et al. regarding variance-driven

adaptation after-effects, which suggest that variance is an abstract property that works independently from its sensory origin and generalizes across domains (Payzan-LeNestour et al., 2016). Michael and colleagues (Michael et al., 2014) also propose variance as an independent property from ensemble average, but suggest that, regarding priming, it operates through feature-specific channels. In our experiments we used a single formalization of variance – dispersion of a dot-motion cloud - so the degree to which our results will generalise to other variance-related serial dependencies requires further investigation. At any rate, it is possible that serial dependence arises at different levels of the hierarchy for diverse feature-dimensions, depending on the areas responsible for the processing of each dimension.

Considering this, we decided to explore the influence of high-level processing on serial dependence in a well-studied low-level feature such as orientation, which is mainly processed at the primary visual cortex (V1) and has been used extensively in serial dependence studies. With this aim we set up a serial dependence experiment wherein half of the trials were masked by continuous flash suppression (CFS). CFS masking of a monocular stimulus is able to completely block visual awareness while preserving V1-based sensory processing, as demonstrated by the persistence of neural adaptation (albeit weakened) in relation to masked, unseen stimuli (Blake et al., 2006; Lin & He, 2009). Our results in Experiment 4 show that serial dependence in relation to the orientation of n-1 unmasked, monocular Gabors exhibits a similar pattern (in terms of tuning width and amplitude of the effect) to that described in previous studies (Fischer & Whitney, 2014; Fritsche et al., 2017). On the contrary, when the previous (n-1) Gabor patch was masked by continuous flash suppression and reportedly unseen by participants (score 1 in the Perceptual Awareness Scale), serial dependence was not detected. Superficially, this result may seem similar to the evidence for a role of attention in gating serial dependence (Fischer & Whitney, 2014). However, it has been shown that spatial attention is able to modulate another history-dependent bias such as adaptation to oriented patterns (a V1-based sensory bias), even in absence of awareness (Bahrami, Carmel, Walsh, Rees, & Lavie, 2008). Furthermore, tilt adaptation

is attenuated, but still present, for unattended and/or invisible stimuli (He & MacLeod, 2001; Spivey & Spirn, 2000). By contrast, our findings indicate that, unlike adaptation, serial dependence cannot arise simply by lingering local changes in sensory areas, but requires higher-level visual processing of the past stimulus (i.e. the inducer).

What are the perceptual/neural mechanisms underlying the observed positive serial dependence? Although this is still uncertain, previous works have proposed exposure-related gain changes or shifts in the neural tuning (Fischer & Whitney, 2014)). Furthermore, its behaviour resembles that implied by Bayesian frameworks of information processing, in which judgments about a certain dimension are attracted towards prior information. Several studies have recognised that the observed systematic errors in magnitude estimation tasks, across diverse dimensions can be well accounted for by assuming an iteratively updated prior, in which recent information is given more weight compared to the overall statistical properties of the environment (Cicchini et al., 2014; Luca & Rhodes, 2016; Petzschner & Glasauer, 2011; Roach et al., 2017). As previously mentioned, the dimension-specific time and capacity limitations encountered in our experiments suggest memory-based operations wherein current decisional representations are merged with previous representations to produce both a contextually-informed decision and the basis for a new prior to inform subsequent trials, in a process akin to a Kalman filter. Variance-related positive serial dependence indeed shares many characteristics with recursive Bayesian dynamics, including the greater weight of more recent information and the association with high confidence in past trials. Positive serial dependence is probably Bayesian-like in many aspects, but there are some nuances to perceptual decision-making that demand further investigation.

Besides the positive serial dependence by recent decisions, in our experiments we also found a longer-lasting negative bias in relation to less recent variance presentations. The basis of this bias is less conclusive, but may be related to adaptation after-effects, like the variance adaptation described in (Payzan-LeNestour et al., 2016). The fact that the negative effect is observed in relation with individual presentations lasting only 500 ms,

appears as early as for the following trial, and remains even for trial n-9 could seem unusual for a 'sensory' after-effect. However, negative after-effects in response to sub-second stimuli have been described previously ((Ryotal Kanai & Verstraten, 2005), (Fritsche et al., 2017)), and in (Fritsche et al., 2017) it also lasts for several seconds. In (Fritsche et al., 2017), the authors propose that it is not the stimulus itself, but a memory trace that causes the negative after-effect on orientation. It is likely that the observed relationship between the current trial and a specific trial in history (e.g. n-5) is actually driven by a broader, averaged contextual representation and not by the individual stimuli several trials removed from the present.

Although our experiments were not designed to ascertain the characteristics of this negative effect, some of its properties could be pinpointed. Our control analyses ruled out that it was a statistical artefact intrinsic to the trial structure of our experiments, since it does not appear in relation to future presentations or shuffled responses, while it persists in random subsets (that are no longer closed sets as the experimental blocks as a whole). The scope of its influence seems to be dependent on time-discount, as it peaks and declines at more recent trial positions (compared to the current trial) in Experiment 3 compared to Experiment 1, likely related to the longer inter-trial intervals in Experiment 3. The amplitude of its effect may also depend on the duration of the RDK stimulus (as would be natural in exposure-dependent adaptation effects), since the negative effect seems weaker in those Experiments where we used 250-ms RDKs instead of 500 ms - however, other changes in the experimental structure might have contributed to this weakening. Intriguingly, some aspects of this negative bias could point to a decisional component, including its independence of retinal location and predominance in low-confidence trials. In any case, the line between perceptual and post-perceptual after-effects may be blurred concerning high-level dimensions such as statistical properties (Payzan-LeNestour et al., 2016; Storrs, 2015).

Some previous studies on different features, both low-level (namely motion (Ryotal Kanai & Verstraten, 2005) and orientation (Fritsche et al., 2017)) as well as high-level

(such as face attributes (Taubert et al., 2016)) have reported concomitant positive and negative biases exerted by the same stimulus, similar to our own results. In Kanai et al. (Ryotal Kanai & Verstraten, 2005), a negative rapid motion after-effect of sensory origin (rMAE) was elicited by a short, sub-second sine-wave luminance grating presented immediately before. However, when the inter-stimulus interval (ISI) was long enough (>3 seconds), a positive bias was elicited instead, in response to the percept and not the low-level sensory signal (as proven by the use of ambiguous motion adaptors). In Fritsche et al. (45), opposite effects of recent history on orientation judgments were exerted by perception (negative bias) and decision (positive serial dependence), very much in line with our findings. The authors proposed that each of these effects has a different biological function, namely increasing sensitivity to changes within the current sensory context and promoting perceptual stability by smoothing away transient variations and noise. Taubert et al. suggest the same duality in their study of serial dependences in face attributes (Taubert et al., 2016), although in their case positive and negative biases are exerted concomitantly by different high-level features of the same visual stimulus (faces): stable traits such as gender would be subject to positive biases in order to smooth away noise, whereas negative after-effects maximizing sensitivity would predominate in changeable attributes such as facial expression.

Given the properties of the positive and negative serial dependencies encountered in our data, and our conclusions about their likely mechanisms, we built a two-layer model for perceptual decision-making, where the two history-dependent biases were produced at different levels of processing, operationalized by a similar approach as previous studies had taken to model each bias separately: (negative) sensory adaptation by population codes and recency bias (or regression effects) by Bayesian Kalman filters. In each iteration of the model, a stimulus (RDK variance) was transduced into a probabilistic sensory response by neural population codes subject to exposure-dependent gain changes -responsible for negative adaptation-, while a decision layer combined the sensory response with prior information to produce a Bayesian judgment about the stimulus -thus producing attractive serial dependence. Such a model was able

to reproduce the observed reversal from positive and negative history-dependent biases in their approximate timescales.

In summary, our study on visual variance reveals two opposite inter-trial dependences that operate at different timescales and likely arise at different levels of perceptual decision-making: a positive serial dependence in relation to high-confidence, dimension-specific decisions and subject to capacity limitations, and a longer lasting negative bias of likely sensory origin. Further investigations are needed to elucidate the precise mechanistic basis of variance-related serial dependence, whether it generalizes to other instances of variance and its relationship to other instances of serial dependence, as well as to explore a possible role of serial dependence in variance in tuning the processing of other visual features to the current levels of external and internal noise, to handle the reliability of sensory information for the construction of a precise yet stable visual experience.

PART III: Contextual Modulation Across the Visual Field: Sensory Processing Under the Uniformity Illusion

Abstract

Part III presents a series of studies concerning contextual modulation of visual variability across the visual field. Specifically, we examine the mechanisms of fovea-periphery integration and the construction of a uniform perceptual experience across various eccentricities by studying the recently described Uniformity Illusion (UI), wherein a pattern with different characteristics in fovea and periphery takes the uniform appearance of the foveal region - dismissing visual variability presented in the peripheral field. By generating adaptation to a pattern suitable for producing UI, we ascertained whether the illusion worked by changing sensory encoding in peripheral receptive fields -in which case we should observe adaptation by the illusory, not the physical properties. In Experiment 5 we applied this paradigm to orientation, and found that the tilt after-effect only ever followed the physical, not the illusory orientation, even for long times of reported illusory experience. This indicated that UI on orientation did not depend on V1 activity patterns. We then tested UI on texture-density, a dimension that is processed beyond V1 (Experiment 6): while potential UI-driven adaptation was not necessary to explain the response bias driven by different adapting patterns, we could not completely rule out such effect due to lack of spatial and temporal specificity of density adaptation, even by the physical patterns. In conclusion, the mechanisms of foveal-peripheral integration and management of visual variability across the visual field, as exemplified by UI, do not require reconstruction of sensory information in the primary visual cortex; it is possible that UI arises as a result of a combination of texture processing and perceptual inflation in the periphery. The contents of Chapter 7 have been published in i-Perception journal under the title 'The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence from Sensory Adaptation' (Suárez-Pinilla, Seth, & Roseboom, 2018a) .

CHAPTER 7: EXPERIMENT 5: INSIGHTS ON THE MECHANISTIC BASIS OF THE UNIFORMITY ILLUSION IN ORIENTATION. EVIDENCE FROM SENSORY ADAPTATION

Visual experience appears richly detailed despite the poor resolution of the majority of the visual field, thanks to foveal-peripheral integration. The recently described Uniformity Illusion (UI), wherein peripheral elements of a pattern take on the appearance of foveal elements, may shed light on this integration. We examined the basis of UI by generating adaptation to a pattern of Gabors suitable for producing UI on orientation. After removing the pattern, participants reported the tilt of a single peripheral Gabor. The tilt after-effect followed the physical adapting orientation rather than the global orientation perceived under UI, even when the illusion had been reported for a long time. Conversely, a control experiment replacing illusory uniformity with a physically uniform Gabor pattern for the same durations did produce an after-effect to the global orientation. Results indicate that UI is not associated with changes in sensory encoding at V1, but likely depends on higher-level processes. The contents of this chapter have been published in i-Perception journal under the title 'The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence from Sensory Adaptation' (Suárez-Pinilla et al., 2018a) .

1. INTRODUCTION

Visual experience appears richly detailed despite the poor sensory precision of the vast majority of the visual field - the visual periphery, extending from roughly 2° to nearly 180° of eccentricity in the horizontal diameter (Strasburger, Rentschler, & Jüttner, 2011). This topic has received considerable recent attention (Michael A. Cohen et al., 2016; Haun, Tononi, Koch, & Tsuchiya, 2017b), with debate about the degree to which visual experience is in fact rich, or rather an inflated illusion not sustained on actual content, as well as about the potential perceptual processes that may contribute to apparent richness. One recent study demonstrated a compelling example of how the rich detail within the high-precision central visual field alters peripheral perception - the Uniformity Illusion (UI) (Otten et al., 2016). UI describes a phenomenon wherein apparent perceptual uniformity occurs when variable sensory stimulation is presented in the peripheral vision, while the central visual field is presented with uniform (or low-variance) stimuli. UI occurs for a wide variety of perceptual dimensions, including relatively low-level sensory features like orientation or colour, and higher-level features such as density (see www.uniformillusion.com for examples). This prompts the intuition that UI may describe a widespread phenomenon that promotes a uniform and detailed visual experience across the entirety of the visual field. But the specific mechanisms and brain regions that give rise to this illusion are still unknown.

We sought to examine the mechanisms underlying UI using perceptual adaptation. It is well established that exposure to a specific stimulus magnitude (like an oriented grating) causes perceptual after-effects (e.g. tilt after-effect; TAE) (Gibson & Radner, 1937). For visual orientation, perceptual after-effects have been associated with specific changes in neural coding at the primary visual cortex (V1) and are localised in a retinotopic reference frame (Knapen, Rolfs, Wexler, & Cavanagh, 2010). In this experiment, we

utilised the spatial specificity of TAE to examine whether the apparent perceptual uniformity in UI can be attributed to changes in V1-based neural coding for visual orientation. Specifically, we presented participants with Gabor grids wherein the orientation of central elements was uniform, but the orientation of peripheral elements was variable - producing UI. At fixed test locations in the periphery of the grid, we presented a physical orientation that differed from the global illusory percept, thus putting local and global orientation in opposition. Following prolonged exposure to global illusory uniformity (UI), we contrasted whether the resultant TAE was consistent with the local, physical orientation or the illusory global orientation.

2. METHODS

2.1. PROCEDURE

The experiment had two parts: Illusion session and Control session. Each session contained six blocks, and each block had an adaptation phase and a test phase (Figure 18). A practice block was run before the Illusion session to familiarise participants with UI.

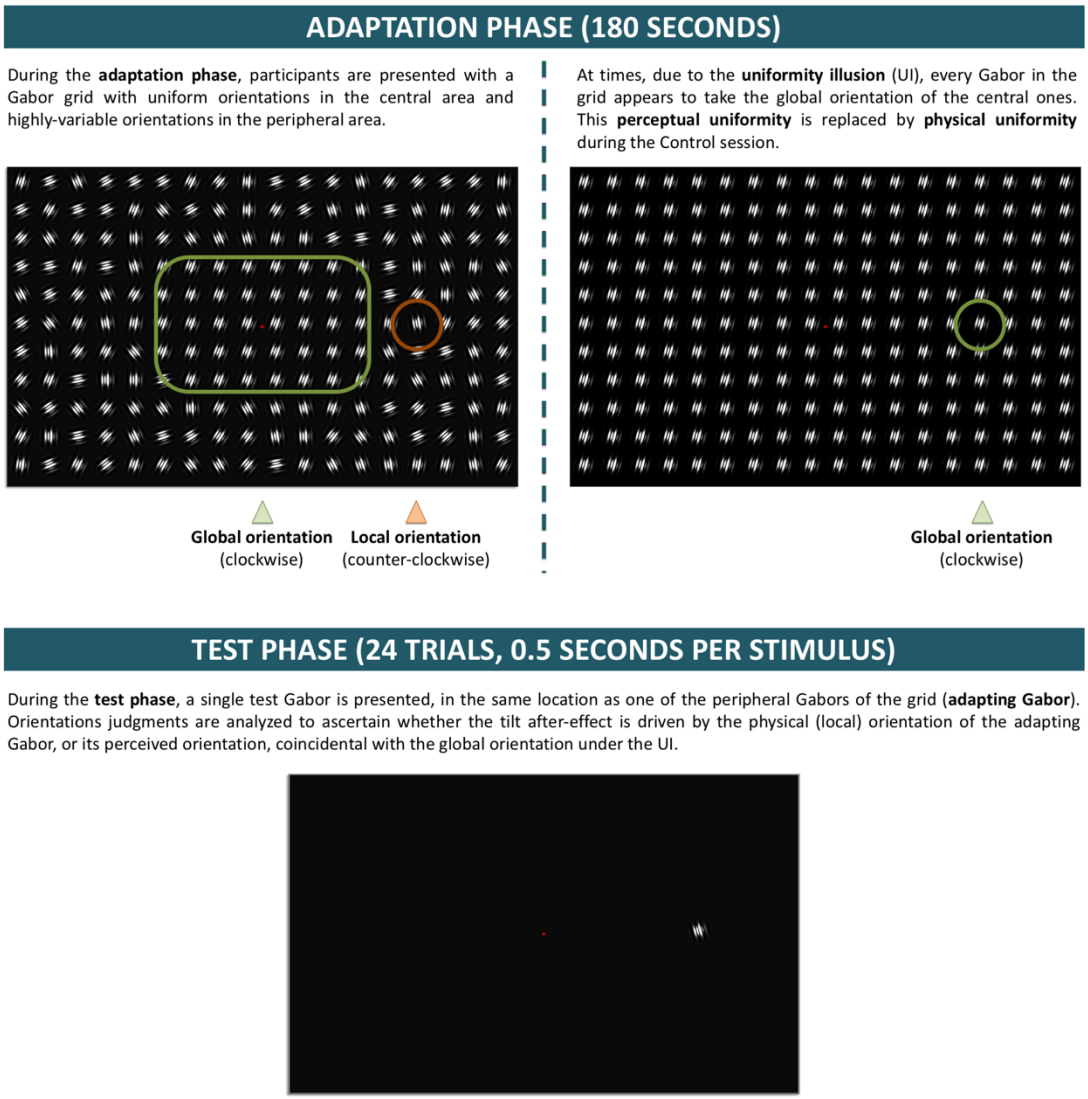


Figure 18. Experiment 5: Methods. During the adaptation phase, participants were presented with a Gabor grid wherein the central Gabors had a uniform orientation, while peripheral orientations were heterogeneous. Under UI, perceptual experience was that of a uniform pattern with all Gabors tilted like the central ones. This illusory percept alternated with a non-illusory, non-uniform percept at different times during adaptation. For a specific peripheral Gabor (adapting Gabor), physical and illusory orientation were always in opposition. The Control session replicated the phenomenology of the Illusion session, replacing perceived with physical uniformity at times in which the participant reported UI in the Illusion session. The test phase had 24 trials, wherein participants reported the tilt of a single peripheral Gabor whose location coincided with the adapting Gabor.

2.1.1. Illusion session

Each block began with an adaptation phase, in which participants were presented with a grid of Gabor patches suitable for producing the UI, affecting the apparent orientation of peripheral elements: all Gabors in the central area had a uniform orientation, whereas orientation of the peripheral Gabors was heterogeneous. Gaze-contingent stimulus presentation ensured that each Gabor was presented to a specific retinal location, as the entire pattern was removed if the participant's gaze deviated from central fixation by more than 1.5 degrees of visual angle (dva) –a tolerance threshold equivalent to half the size of each cell of the grid. Adaptation lasted 180 seconds but, because the stimulus was removed when fixation lapsed, actual exposure time could be shorter.

Participants reported the experience of illusory uniformity by pressing a key when all Gabors appeared to take a uniform orientation.

The test phase had 24 trials, separated by a pseudo-random interval of 1000-1500 ms. In each trial, a single Gabor (test Gabor) was presented for 500 ms at a specific peripheral location, coinciding with the position of a specific Gabor during adaptation (adapting Gabor). Participants reported if the test Gabor was tilted clockwise (CW) or counter-clockwise (XCW) from vertical.

2.1.2. Control session

The Control session also had six blocks, each built to replicate the phenomenology of a homologous block of the Illusion session but replacing illusory for physical uniformity during the adaptation phase.

During the adaptation phase in the Illusion session, an empty background was presented whenever the gaze-contingent mechanism removed the adapting pattern. The same pattern of stimulus presentation and removal was replicated in the Control session. The stimulus was additionally removed whenever fixation lapsed in the Control session. At any other time, the presentation displayed one of two patterns, differing only in the orientation of peripheral Gabors. The first was identical to the pattern presented in the Illusion session and was displayed at times in which the participant had *not* reported UI during adaptation in the Illusion session. At times during which the participant had reported UI, the presented pattern was one in which all Gabors had the same *physical* orientation, consistent with the desired illusory orientation during the Illusion session. Thus, physical uniformity was inserted at the times in which illusory uniformity had been reported in the Illusion session. Participants were not informed that this would occur.

The test phase was identical to that in the Illusion session: the location and orientation of the test Gabor in each trial was identical, as well as its test latency (time between the end of the adaptation phase and stimulus onset).

2.2. STIMULI

Stimuli were displayed on dark grey background (1.96 cd/m^2). A red fixation dot (8.34 cd/m^2, 0.42 dva diameter) showed constantly on the screen centre.

2.2.1. Gabor patches

Each Gabor consisted of a sine-wave luminance grating with Michelson contrast of 1, 0º phase and spatial frequency of 1.66 cycles per dva (cpd), and a 2-D Gaussian envelope with a sigma of 0.43 dva.

2.2.2. Adapting pattern

The adapting pattern spanned the entire screen and consisted of a 13x17 grid formed by invisible square cells measuring 3 dva per side (Figure 18). Each Gabor was presented in the centre of each cell. The central area spanned 15 dva horizontally and vertically, encompassing all cells belonging to rows 5-9 and columns 7-11. All central Gabors had the same orientation, which could be one of two values, each for half the blocks of one session: -15^{o} (global clockwise tilt, GCW) or 15^{o} (global counter-clockwise tilt, GXCW). The orientations of peripheral Gabors were sampled from a discrete uniform distribution centred on the global orientation and ranging 70^{o} (35^{o} to each side). Thus, mean orientation was the same for central and peripheral Gabors and matched the global orientation perceived under UI.

Two peripheral Gabors of the pattern (adapting Gabors) corresponded to the positions in which the test Gabors would be displayed during the test phase: they were located along the middle (7th) row, at 12.02 dva left and right of the screen centre (columns 5 and 13). Both had the same non-randomized local orientation, which was the opposite of the global orientation of the block: either 15^{o} (local counter-clockwise tilt, LXCW) or -15^{o} (local clock-wise, LCW).

Henceforth we give the label **adapting condition CX** to the presentation pattern wherein the local orientation of the adapting Gabor is clockwise and the global orientation of the pattern is counter-clockwise (LCW, GXCW). Conversely, we will refer to the pattern with LXCW and GCW orientations as **adapting condition XC**. Both conditions occurred equally frequently during the experiment.

As described above, during the Control session, the adapting pattern was replaced by a physically uniform pattern at those times during which participants had reported UI in the Illusion session. In these instances, *every* Gabor in the pattern (including the adapting Gabors) took the global orientation.

2.2.3. Test Gabors

A single test Gabor was presented per trial, matching the position of one of the two adapting Gabors. Test Gabors were displayed in the left and right hemifield with equal frequency per block and could take one of eight equally frequent orientations: -12°, -5°, -2°, -1°, 1°, 2°, 5° and 12° (negative values indicate clockwise tilt). Thus, test orientations were always intermediate between global and local orientations (-15°, 15°).

2.4. PARTICIPANTS

Participants were recruited through online advertisement, over 18 and reported normal or corrected-to-normal vision. This study received ethical approval by the Research Ethics Committee of the University of Sussex.

2.5. APPARATUS

Experiments were programmed in MATLAB 2016a (MathWorks Inc., Natick, US-MA) and displayed on a LaCie Electron 22BLUE II 22'' with screen resolution of 1024x768 pixels and refresh rate of 100 Hz. Eye-tracking was performed with Eyelink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at sampling rate of 1000 Hz, with level desktop camera mount. Head position was stabilized 43 cm from the screen using chin and forehead rest. Calibration of the eye-tracker was performed at the beginning of each block with a standard five-point grid and a maximal average error of 0.5 dva.

2.6. STATISTICAL ANALYSIS

Psychometric curve fitting was performed in MATLAB 2017b, using Palamedes toolbox, version 1.8.1 (Prins & Kingdom, 2009). A cumulative Gaussian curve was fitted by the method of maximum likelihood to the proportion of 'counter-clockwise' (XCW) responses per test Gabor orientation, separately for each participant and session/condition (depending on the specific analysis). The threshold (α) for 0.5 proportion of XCW responses and the slope of the curve (β) were free parameters (starting values: $\alpha=0^{o}$, $\beta=0.04$), while guessing (γ) and lapse rate (λ) were fixed at zero.

Bayesian statistics were conducted on JASP (JASP Team (2017), version 0.8.3.1). For Bayesian t-tests we employed as prior distribution Cauchy(0, $\frac{1}{2}\sqrt{2}$) for two-sided predictions, or a folded distribution for one-sided predictions: Cauchy^{+}(0, $\frac{1}{2}\sqrt{2}$) for predictions of the form measure 1>measure 2; Cauchy^{-}(0, $\frac{1}{2}\sqrt{2}$) for the reverse. Likewise, for Bayesian Pearson correlations we employed a uniform prior U(-1,1) for two-sided predictions, or U(-1,0)/U(0,1) for one-sided (negative/positive) predictions, respectively. For each contrast result, the prior utilised is indicated by the formulated prediction and the subscripts in BF_{10} (two-sided) or BF_{-0} /BF_{+0} (one-sided).

3. RESULTS

Thirty participants volunteered for the experiment: 23 female, mean age 21.6.

To ensure sufficient exposure to the adapting pattern, we excluded blocks wherein the pattern had been displayed for less than 2/3 of the adaptation phase (<120 seconds), due to gaze-contingent stimulus removal. In such cases, the corresponding blocks from

both Control and Illusion sessions were removed, to maintain balance. This caused exclusion of 32.78% blocks (118/360), including the entire datasets from five participants. Results presented here correspond to the blocks of the remaining twenty-five participants. Furthermore, since our analyses compared responses across adapting conditions (CX/XC), two additional participants were excluded as all their valid blocks were of only one condition. Results presented here correspond to the remaining 23 participants. Overall results for all 25 participants with valid blocks were very similar to this counterbalanced sample (see section 4 of the current chapter, 'Supplementary Analyses').

3.1. ADAPTATION PHASE

Average exposure time to the adapting pattern per block was 164.13 and 149.47 seconds for the Illusion and Control sessions: 91.18% and 83.04% of the adaptation phase, respectively. The lower proportion in the Control session was expected as pattern removal occurred whenever it had in the Illusion session, in addition to times of improper fixation in the Control block.

Perceived uniformity was reported, on average, for 43.48 seconds in the Illusion session, 26.77% of the time of pattern presentation (minimum 0.55%, maximum 72.23%). The proportion of time of perceived uniformity during the Control session was similar to that for the Illusion session: 28.41% (minimum 0.59%, maximum 78.42%, Bayesian paired-samples t-test: BF_{01}=2.733 - anecdotal evidence for the null hypothesis). Physical uniformity in the Control session was reported as perceptually uniform 68.13% of the time; by contrast, the non-uniform pattern was reported as uniform only 9.24% of the time. It is likely that presentation of a truly uniform pattern at times shifted a subjective criterion for uniformity by comparison, leading to more conservative reports in the Control sessions.

3.2. HYPOTHESES AND MEASUREMENTS

The experiment placed adaptation to illusory and physical orientation in opposition to disambiguate between two competing hypotheses:

1. The perceived orientation under UI has no effect on tilt adaptation; the TAE is driven solely by the physical orientation of the adapting Gabor.

2. The global orientation perceived for the entire pattern (including the adapting Gabor) under UI can produce a TAE.

To decide between hypotheses, data was analysed to ascertain the direction of the adaptation-induced bias. During the Illusion session, a TAE driven by (i.e. away from) the local orientation of the adapting Gabor would imply physical adaptation, while a global-driven TAE would indicate adaptation to illusory orientation. During the Control session, both local and global-driven TAE are compatible with physical adaptation, since the adapting Gabor physically takes the global orientation at times of reported illusory uniformity in the Illusion session.

For each session, we calculated the proportion of XCW reports per test Gabor orientation and obtained two types of summary measures for the direction of the bias observed in each given condition: response-based measures (directly based on the proportion of reports) and Psychometric function-based measures (based on Psychometric curve fitting, as the name indicates). Subsequently we ascertained whether these measures were dependent on the conditions of the adaptation phase, and if so, which hypothesis was consistent with such dependency.

3.2.1. Response-based measures

We defined the variable %XCW as the proportion of 'XCW' reports per participant, session and adapting condition (CX/XC). A participant may report either 'CW' (clockwise) or 'XCW' (counter-clockwise) for every presented test Gabor. Note that test Gabor CW and XCW orientations are equally frequent per block, with an average orientation of 0^o for all sessions and adapting conditions. Therefore, in absence of any response bias we should expect %XCW=50%. A lower proportion (%XCW<50%) would indicate a CW bias; conversely, %XCW>50% would reflect a XCW bias.

Therefore, %XCW indicates whether responses are biased to the CW or XCW orientation. However, for summarising responses across different conditions (CX, XC), we needed to define more abstract measures that expressed whether the TAE was driven by (i.e. away from) the local or the global orientation of the adapting pattern. Thus, we defined %Local as the proportion of responses in the same direction as the 'local' orientation of the block: i.e., the proportion of CW reports in CX adapting condition (100-%XCW), and the proportion of XCW reports (%XCW) in XC condition. A %Local<50% would indicate a local-driven TAE (responses are biased away from the local orientation), and %Local>50% would reflect a global-driven TAE.

3.2.2. Psychometric function-based measures

As described in the 'Methods' section (section 2 of the current chapter), we obtained the best-fitting cumulative Gaussian psychometric curve on the proportion of responses per participant, session and condition, and defined the point of subjective equality (PSE) as the test orientation at which 50% reports are XCW. Since CW orientations have (conventionally) negative sign in our experiment and vice versa, negative PSE indicates a XCW bias and positive PSE a CW bias. An unbiased response pattern would correspond to PSE=0^o. Similar to response-based approach, we obtained two measures:

1. **PSE_{CX}** and **PSE_{XC}**, i.e. the PSE corresponding to adapting condition CX and XC, respectively. These measures present a negative correlation with their equivalent response-based measures: %XCW<50% and PSE>0^{o} both indicate a CW bias, while %XCW>50% and PSE<0^{o} reflect a XCW bias.

2. **dPSE**=PSE_{CX}–PSE_{XC}. As with %Local, we employ dPSE as a summary measure indicating the overall direction of the bias. A dPSE<0 would indicate a local-driven TAE (PSE_{CX}<PSE_{XC}) and a dPSE>0 a global-driven TAE (PSE_{CX}>PSE_{XC}). Therefore, %Local and dPSE measures are positively correlated within participant, session and condition.

In summary, for a local-driven TAE, responses for adapting condition CX should exhibit a XCW bias compared to condition XC: $\%XCW_{CX}$>$\%XCW_{XC}$, %Local<50%, PSE_{CX}<PSE_{XC}, dPSE<0, consistent with physical adaptation to the local orientation. The reverse should happen for a global-driven TAE: $\%XCW_{CX}$<$\%XCW_{XC}$, %Local>50%, PSE_{CX}>PSE_{XC}, dPSE>0 are consistent with adaptation to the illusion (or to the physical replication of the illusion during the Control session).

3.3.TAE IS DRIVEN BY PHYSICAL, NOT ILLUSORY ORIENTATION

3.3.1. Overall effect

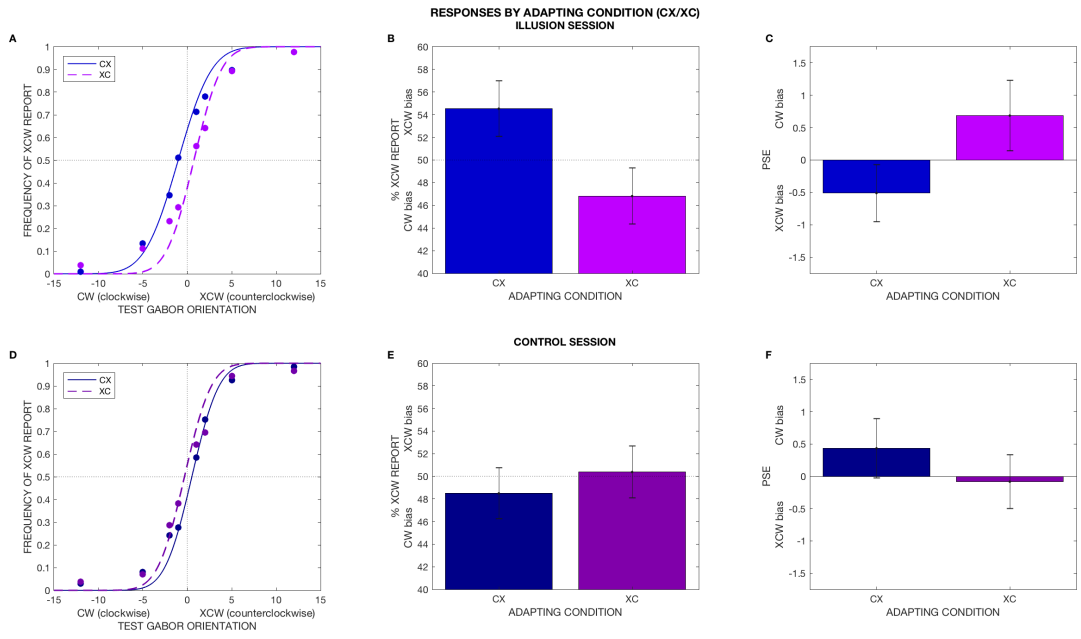


Figure 19. Experiment 5: Overall results. Response patterns by adapting condition. Illusion **(19a-19c)** and Control **(19d-19f)** session. Figures **19a** and **19d** present the sample's proportion of counter-clockwise (XCW) reports per test Gabor orientation, separated by adapting condition, during the Illusion (**19a**) and Control (**19d**) sessions. The dotted lines show the best cumulative Gaussian fit for the psychometric curve of each condition, fitted on the sample's pooled data (N=23). These figures are included for illustrative purposes only, as the PSEs obtained for analysis were computed separately for each participant's data -as presented in **19c** and **19f**, see below for details. **19b** and **19e** present the average %XCW computed separately by participant, adapting condition and session. **19c** and **19f** depict the average point of subjective equality (PSE) computed separately per participant and condition. In all bar graphs, the bar heights represent the sample's average and the error bars the between-participant standard error. **19b-c.** Illusion session. The %XCW and PSEs for both adapting conditions reflect a bias away from local orientation (local-driven TAE). **19e-f.** Control session. On average, responses show a global-driven TAE in CX condition and are unbiased in XC. These results show that perceived (illusion) and physical (control) uniformity behave differently, suggesting that the TAE is always driven by the physical orientation, even when that orientation is unseen under UI.

Illusion session

Figure 19a presents the sample's average proportion of XCW reports per test Gabor orientation during the Illusion session, separated by adapting condition (CX or XC). For illustration purposes, it shows cumulative Gaussian curves fitted for the sample's pooled data (N=23). However, for data analysis (PSE) we fitted a Psychometric function on each participant's responses separately: individual fits are detailed in the section 4 of the current chapter, 'Supplementary Analyses'. Individual PSEs for each adapting condition are summarized in figure 19c. Likewise, participants' proportion of XCW reports (%XCW) per adapting condition are summarized in figure 19b.

On average, $\%XCW_{CX}$=54.533% and $\%XCW_{XC}$=46.818% reflected a XCW and CW bias, respectively ($\%XCW_{CX}>\%XCW_{XC}$ Bayesian paired-samples t-test: BF_{0+}=7.244). In other words, reports were biased away from the local orientation in both conditions (%Local =46.245%).

Psychometric curve analysis showed converging results: PSE_{CX}=-0.502º, PSE_{XC}=0.687º, again indicating a XCW and CW bias, respectively: dPSE=-1.197º ($PSE_{CX}<PSE_{XC}$ Bayesian paired-samples t-test: BF_{-0}=3.057).

In summary, the response pattern during the Illusion session indicated a local, physical-driven adaptation.

Control session

In the Control session, the adapting Gabor physically takes the global orientation of the pattern during times of reported uniformity in the Illusion session. If adaptation is produced by physical orientation, we should observe a more global-driven TAE compared to the Illusion session: $\%Local_{IL}<\%Local_{CO}$, $dPSE_{IL}<dPSE_{CO}$. Conversely, if perceived orientation under UI causes adaptation, we should not see a difference between perceived and physical uniformity: $\%Local_{IL}=\%Local_{CO}$, $dPSE_{IL}=dPSE_{CO}$.

As shown in figures 19e-19f, results indicate predominance of global-driven TAE during the Control session: specifically, a CW bias for condition CX ($\%XCW_{CX\ CO}$=48.297%, $PSE_{CX\ CO}$=-0.083º), indicative of a global TAE, and lack of any clear bias for condition XC ($\%XCW_{XC\ CO}$=50.362%, $PSE_{XC\ CO}$=0.433º) – the latter likely reflecting a mixture of local and global-driven TAE in relation with the alternation of two physical adapting orientations. The proportion of reports in the direction of each block's local orientation ($\%Local_{CO}$) was 51.007%, compared with 46.245% during the Illusion session. Bayesian t-test showed strong evidence for $\%Local_{IL}$<$\%Local_{CO}$: BF_{-0}=12.862. Psychometric function fitting provided converging results: PSE_{CX-CO}=0.433º, PSE_{XC-CO}=-0.083º, $dPSE_{CO}$=0.516º. A Bayesian paired-samples t-test comparing dPSE in both sessions was consistent with physical-driven adaptation: $dPSE_{IL}$<$dPSE_{CO}$, BF_{-0}=7.476. Therefore, the absence of a global-driven TAE in the Illusion session was not simply because the global pattern of orientation was insufficient to induce TAE – rather, the illusory (but not the physical) global pattern was insufficient to induce TAE.

The overall predominance of global-driven TAE in the Control session, despite presentation of the uniform pattern for only ~27% of time, may be related to a putatively stronger adaptation during this time due to the adjacent Gabors, which then take the global orientation, contributing to the receptive field(s) where the test Gabor will be later presented. Note, however, that the size of each grid cell (3 dva) is larger than the diameter of most receptive fields at V1 (around 1 dva) (Bentley & Salinas, 2009), and the relationship between stimulus size and TAE strength is not straightforward (Harris & Calvert, 1985; Parker, 1972). Another possibility involves extra-classical receptive field effects exerted by the global surround on the adapting Gabor when the latter takes the global orientation (iso-orientation surround suppression) (Chen, Chen, Gao, Yang, & Yan, 2015). Whatever the contribution of these effects, they act differently on physical compared to illusory iso-orientation, in the manner expected for low-level processing of the former, but not the latter.

3.3.2. Time-dependent effect

Overall, responses in the Illusion and Control session fit the hypothesis that TAE under UI is only driven by physical, and *not* illusory orientation. However, in the Illusion session UI is perceived during only ~27% of pattern exposure, on average. Thus, it could be argued that a global, illusion-driven TAE might have been present, but undetected in the overall results -overshadowed by the local-driven TAE at times when UI is not perceived. This possibility seems unlikely, because responses in the Control session (with uniformity also presented ~27% of time) *do* show an influence of the global-driven TAE. Thus, such a possibility could only hold if the TAE driven by illusory orientation was weaker than that caused by physical orientation. To rule out this possibility, we examined the data from the Illusion sessions for evidence of exposure time-dependency of the TAE magnitude. Since the TAE is time-dependent (Patterson, Wissig, & Kohn, 2013), if illusion-driven adaptation was in fact present, we should find evidence for a shift toward more global/less local TAE with longer times of perceived uniformity.

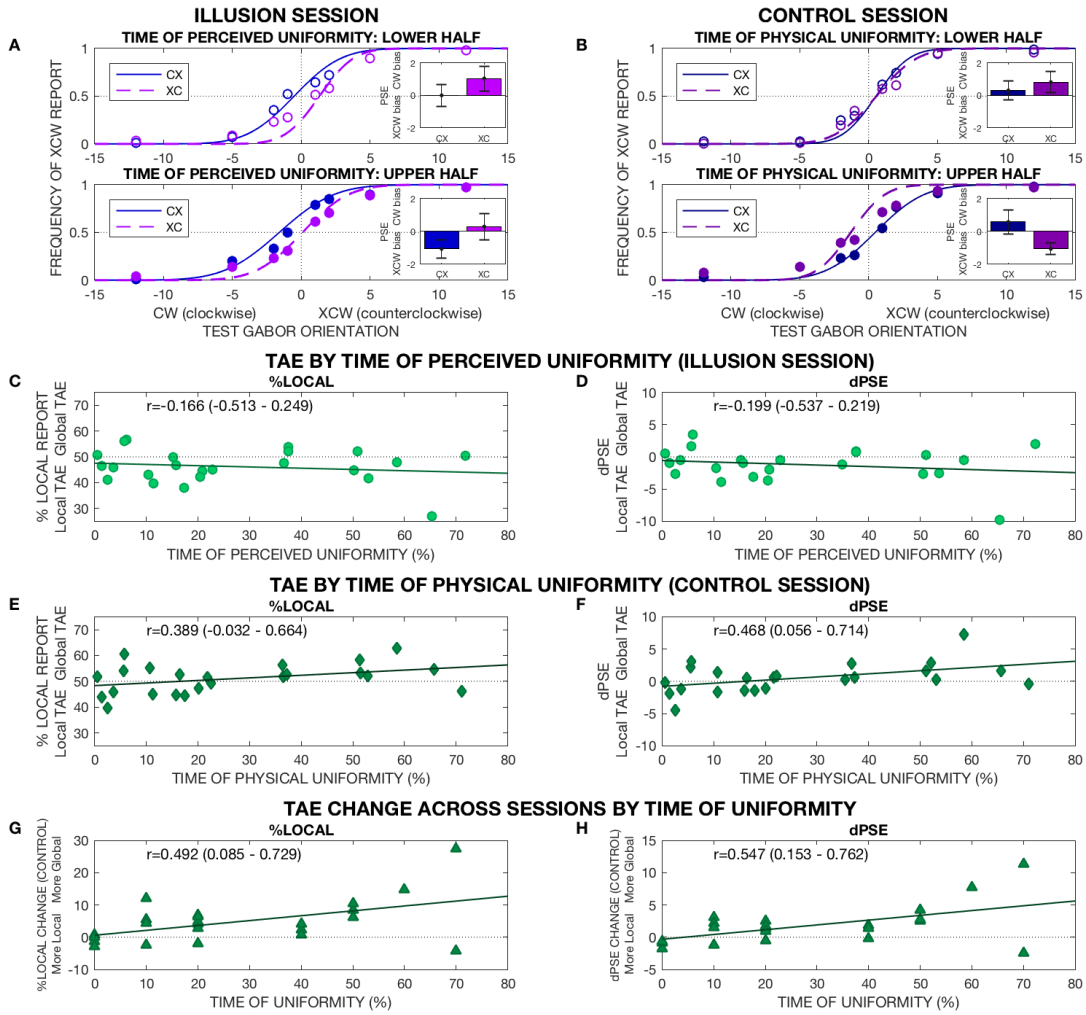


Figure 20. Experiment 5: Time-dependent results. TAE by time of uniformity: physical, but not perceived uniformity, causes a shift toward global-driven TAE in a time-dependent manner. **20a-b** classify participants into two groups according to whether their average time of uniformity is below (lower half) or above (upper half) the sample's median and depict each group's average responses by adapting condition in the Illusion (**20a**) and Control (**20b**) session. For illustration purposes, a Psychometric function fitted to the pooled data is shown in the main figures; however, all analyses are based on Psychometric functions fitted to each participant's data -the group average PSEs of these functions are shown in the insets. In the Illusion session, PSEs indicate local-driven TAE regardless of time of perceived uniformity (except for condition CX in the lower-half group, which shows no noticeable TAE overall). In the Control session the TAE shifts to global-driven for longer presented physical uniformity. **20c-20d** present the correlation between time (%) of perceived uniformity and overall direction of the TAE (local/global driven) during the Illusion session, expressed in terms of %Local (**20c**) and dPSE (**20d**). Each data point represents a participant's average data during the session. Pearson's correlation coefficient and 95% credible intervals are in each panel, showing that there is no significant time-dependency for perceived uniformity. **20e-20f**: Correlation between time (%) of physical uniformity and overall direction of the TAE (local/global driven) during the Control session, expressed in terms of %Local (**20e**) and dPSE (**20f**). Each data point represents a participant's average data during the session. Pearson's

correlation coefficient and 95% credible intervals are in each panel, showing a positive time-dependency of physical uniformity in the expected direction: presentation of the global panel for a longer time leads to more global-driven TAE, unlike the lack of effect of illusory uniformity during the Illusion session. **20g-20h.** TAE change by time of uniformity. The horizontal axis represents each participant's time (%) of Uniformity. For illustrative purposes, it has been rounded to the nearest 10%, so that the horizontal axis is applicable to both sessions of each participant (since physical uniformity is presented for the same time than the duration of the illusion reported in the previous session, but in practice the gaze-contingent stimulus removal during the Control session generates a slight jitter). The vertical axis represents the change in the summary measure of the direction of the TAE (%Local in **20g**, dPSE in **20h**) in the Control session with respect to the Illusion session: i.e., in **20g**, the vertical axis represents $\%Local_{Control}$ - $\%Local_{Illusion}$ and in **20h** it represents $dPSE_{Control}$ - $dPSE_{Illusion.}$ A positive difference indicates that a participant's responses during the Control session are subject to a global-driven bias (higher %Local/dPSE) compared to the Illusion session. Both sessions are identical for each participant, except that uniformity is illusory during the Illusion session, and physical during the Control session. We observe that, for short times of uniformity, a participant's response pattern is almost identical in both sessions (roughly 'zero' difference). However, as time of uniformity increases, the patterns become more different, with the Control session showing a more global-driven bias. Pearson's correlation coefficient confirms this positive correlation, with a BF_{+0}=7.458 for time - %Local change and a BF_{+0}=15.997 for time – dPSE change.

Illusion session

If the TAE is driven *only* by physical orientation, in the Illusion session we should expect independence from time of perceived uniformity. Conversely, if the perceived orientation under UI causes adaptation, the response pattern should shift from predominantly local-driven towards more global-driven for longer time of perceived uniformity. We can assess this potential shift by examining %Local and dPSE. As stated above, %Local<50% and negative dPSE indicate predominance of local-driven TAE, whereas %Local>50% and positive dPSE reflect global-driven TAE. Thus, in the presence of illusion-driven adaptation, %Local and dPSE should correlate positively with time of perceived uniformity.

As time measure, we employed the proportion of perceived uniformity (over time of pattern presentation), for conveying the balance between local and (putative) global effects.

We ran a Bayesian bivariate correlation between each participant's average time of perceived uniformity and %Local, showing a Pearson's r=-0.176 (95% credible intervals -0.509, 0.223), with moderate evidence against a positive correlation: BF_{+0}=0.145. See figure 18c.

Likewise, we analysed the bivariate correlation between time of perceived uniformity and dPSE (Figure 18d). Pearson's correlation coefficient and 95% credible intervals were r=-0.199 (-0.537, 0.219), again showing moderate evidence against a positive correlation: BF_{+0}=0.146.

In summary, evidence opposed any positive association between time of perceived uniformity and a trend toward more global-driven TAE, thus opposing predictions formulated for illusion-based adaptation.

Control session

For the Control session we performed analogous analyses as for the Illusion session, but with time of physical instead of perceived uniformity.

Since global uniformity is a physical stimulus in this session, a time-dependent shift from local to global-driven TAE should be expected regardless of the capacity of illusory orientation to induce a TAE. Thus, this Control session acts as a sanity check to rule out that the failure to find time-dependency in the Illusion session was simply due to insufficient exposure to the global pattern - even in the cases of longest time of uniformity.

We performed a Bayesian bivariate correlation between individual average time of physical uniformity and %Local (Figure 18e). Pearson's coefficient was r=0.414, 95%

credible intervals (0.015, 0.672), showing moderate evidence for a positive correlation: BF_{+0}=3.581.

Computing the correlation between time of physical uniformity and dPSE (figure 3D) rendered r=0.468, 95% credible intervals (0.056 – 0.714), with moderate evidence for a positive correlation: BF_{+0}=5.546.

In summary, physical uniformity presented for durations equivalent to the reported illusory uniformity was sufficient to observe a shift towards a global-driven TAE.

3.4. CONCLUSIONS

We aimed to explore the underlying basis of perceptual uniformity in the Uniformity Illusion by using a version of UI with oriented Gabor patches, and found that UI does not produce an orientation adaptation after-effect consistent with the illusory percept. Instead, orientation after-effects only ever followed the (local) physically presented orientation - even in cases where the illusion had been reported for very long times. This was not due to insufficient exposure to the illusory percept, as a physically uniform control pattern presented for the same time did produce an after-effect to the global orientation. Thus, our results suggest that the UI on orientation is not associated to changes in V1 sensory encoding, but arises from higher-level (higher than primary visual cortex) perceptual processes.

4. SUPPLEMENTARY ANALYSES

4.1. PSYCHOMETRIC CURVE FITTING

4.1.1. Individual Psychometric curves

Figure 21 shows the individual responses (proportion of XCW reports by test Gabor orientation) and best-fitting cumulative Gaussian functions obtained for each participant, with their data split by adapting condition (CX and XC) and session (Illusion and Control). Only the 23 participants with valid data on both adapting conditions are presented. The methodology for cumulative Gaussian fitting is detailed in the Methods section of the current chapter (section 2.6 'Statistical analysis').

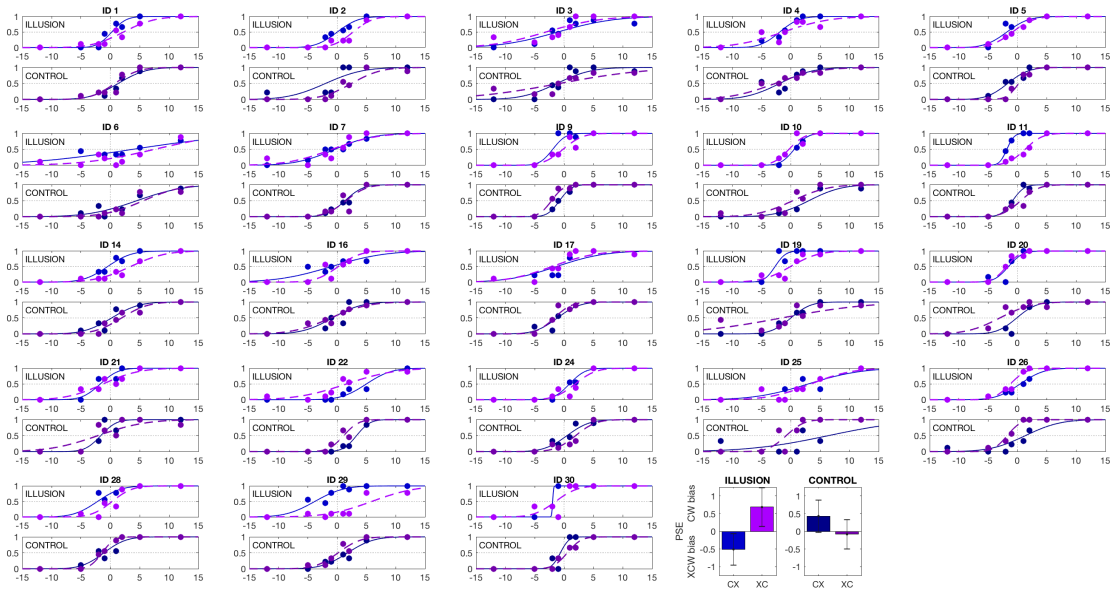


Figure 21. Experiment 5: Individual results. Response pattern and best-fitting Psychometric function for each participant's data, split by adapting condition and session. As indicated in the figure, the upper plot for each participant corresponds to the Illusion session and the lower one to the Control session. In all plots (except for the summary graph at the bottom right), the X axis represents the test Gabor orientation, ranging from -12º to 12º, with

negative and positive sign conventionally indicating CW (clockwise) and XCW (counter-clockwise) tilt, respectively. The Y axis represents the proportion of 'XCW' reports per test Gabor orientation. Blue and purple dots pertain to adapting conditions CX (local orientation CW, global XCW) and XC (local XCW, global CW), respectively. The dotted lines present the best-fitting cumulative Gaussian curve for the participant's data, split by session and adapting condition. A rightward shift of one curve with respect to the other condition (or session) indicates a comparatively more CW bias, and vice versa. In other words, a positive PSE indicates a CW bias and a negative PSE a XCW bias. The summary graph at the bottom right indicates the average of the individual PSEs of all the Psychometric curves shown in the previous plots, separated by session (Illusion session in the left plot and Control session in the right plot) and adapting condition (as labelled in the horizontal axis: CX/XC). The error bars indicate the standard error. For the Illusion session, the negative and positive PSE in conditions CX and XC, respectively, indicate that the tilt after-effect (TAE) is driven by (i.e. away from) the local, physical orientation. For the Control session, we observe, on average, a global-driven TAE for condition CX (PSE>0) and absence of a clear bias in condition XC. These results likely show a 'mixture' of local and global-driven TAE. In the Control session, the adapting Gabor takes the global orientation during the times of presentation of a physically uniform pattern. Thus, results in this session are also compatible with physical-driven adaptation.

4.1.2. Goodness-of-Fit

We assessed the goodness-of-fit separately for each participant and session, by using the function provided to such effect in the Palamedes toolbox (goodness-of-fit test across several conditions, specifically CX and XC conditions). The target model assumed a cumulative Gaussian curve for the relationship between test Gabor orientation and proportion of XCW reports, with varying thresholds (α) and slopes (β) across adapting conditions, and a guess rate (γ) and lapse rate (λ) fixed to zero in both conditions. Table 5 shows the pDev values, indicative of the goodness-of-fit of the target model to the participant's data. The fit was reasonable (pDev$\geq$ 0.05) for most datasets (37/46), and poor only for 9/46. On average, $pDev_{Illusion}$= 0.282 (SE 0.055) and $pDev_{Control}$= 0.213 (SE 0.031) were well above the 0.05 cut-off point; according to a Bayesian paired-samples t-test, there was no evidence for a difference between the goodness-of-fit of the Illusion compared to the Control session (BF_{10}=0.365), with an anecdotal, but tending toward moderate (BF_{10}<=1/3) support for the null hypothesis (i.e. equal goodness-of-fit across sessions).

TABLE 5. Experiment 5: Goodness-of-fit per participant and session

ID	$pDev_{Illusion}$	$pDev_{Control}$
1	0.307	0.328
2	0.515	0.000
3	0.035	0.014
4	0.416	0.320
5	0.372	0.401
6	0.242	0.180
7	0.071	0.247
9	0.007	0.397
10	0.099	0.016
11	0.714	0.337
14	0.866	0.300
16	0.331	0.310
17	0.000	0.182
19	0.079	0.115
20	0.037	0.096
21	0.654	0.037
22	0.172	0.302
24	0.086	0.227
25	0.650	0.092
26	0.565	0.006
28	0.075	0.250
29	0.125	0.539
30	0.071	0.194

Table 5. Experiment 5. Goodness-of-fit for each participant's and session's data. pDev value ranges between 0 and 1: the larger this value, the better the fit.

Thus, the Psychometric function goodness-of-fit to the data is acceptable on average, with only a few individuals who exhibit a poor fit. As detailed below, there is an almost perfectly linear correlation between measures directly based on participants' responses

and measures based on fitted Psychometric functions (PSE), further confirming that the latter measures are accurately tracking each participant's response patterns.

4.1.3. Correlation between individual response-based measurements and Psychometric function-based measurements

As mentioned in section 3.2 of the current chapter ('Hypotheses and measurements'), if the PSE of the fitted Psychometric function is a meaningful measure of each participant's response bias, we should observe a negative correlation between %XCW and PSE: the larger the first, the more responses are biased toward XCW reports, which would be expressed by a lower PSE, and vice versa. Conversely, for the more abstract measures %Local and dPSE, we should observe a positive correlation: the larger these measures, the more responses are biased away from the global orientation (global-driven TAE). Thus, the correlation of response-based and Psychometric function-based measures serves as an indication of the validity of the latter ones. Table 6 and Figure 22 show that this is the case, demonstrating almost perfect negative correlations between %XCW and PSE for all conditions and sessions ($r<-0.95$), as well as almost perfect positive correlations between %Local and dPSE.

TABLE 6. Experiment 5: Correlation between measurement types

6a. %XCW and PSE

Measurements	Pearson's r	95% Credible Interval		BF_{-0}
$\%XCW_{Illusion\ CX}$, $PSE_{Illusion\ CX}$	-0.962	-0.983	-0.896	$1.135*10^{11}$
$\%XCW_{Illusion\ XC}$, $PSE_{Illusion\ XC}$	-0.961	-0.982	-0.892	$7.929*10^{10}$
$\%XCW_{Control\ CX}$, $PSE_{Control\ CX}$	-0.955	-0.980	-0.879	$2.372*10^{10}$
$\%XCW_{Control\ XC}$, $PSE_{Control\ XC}$	-0.973	-0.988	-0.925	$3.349*10^{12}$

6b. %Local and dPSE

Measurements	Pearson's r	95% Credible Interval		BF_{+0}
$\%Local_{Illusion,}$ $dPSE_{Illusion}$	0.955	0.875	0.981	$7.334*10^{9}$
$\%Local_{Control}$, $dPSE_{Control}$	0.946	0.851	0.976	$1.208*10^{9}$

Table 6. Experiment 5. Bivariate correlations between individual response-based and Psychometric function-based measurements. **6a** presents the correlations between %XCW and PSE, computed separately for each session (Illusion/Control) and adapting condition (CX/XC). **6b** presents %Local and dPSE, separately for each session. An almost perfect negative correlation is observed in all cases, indicating that Psychometric function-based measurements are accurately tracking the participant's response bias. BF_{+0} and BF_{-0} refer to the Bayes factor for the existence of a positive or a negative correlation, respectively (extreme evidence in favour of the alternative hypothesis in all cases).

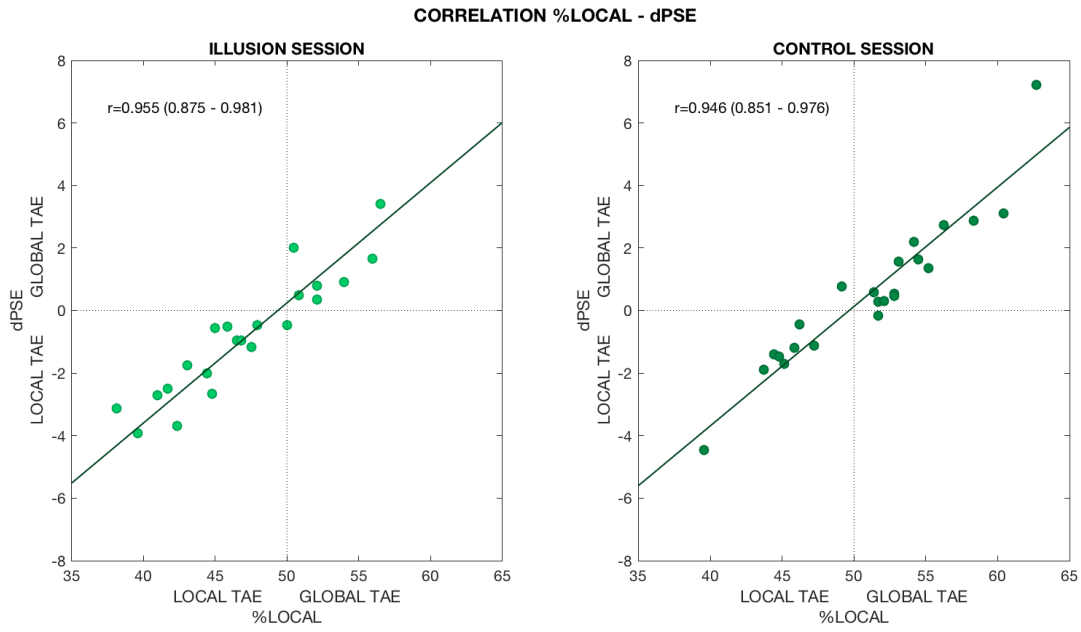


Figure 22. Experiment 5: Measurements. Correlations between individual %Local and dPSE, computed separately for each session (Illusion/Control). An almost perfectly linear positive correlation is observed for both sessions.

4.2. COMPARISON OF RESULTS FOR THE ENTIRE SAMPLE (N=25) AND FOR THE SAMPLE WITH COUNTERBALANCED ADAPTING CONDITIONS (N=23)

Because our analyses involved comparison of response patterns across adapting conditions (CX/XC), we excluded data from two participants who had valid blocks of a single condition. Here we show that the summary results between the entire sample (N=25) and the counterbalanced sample (N=23) are comparable. Table 7a summarizes %XCW and Table 7b the PSE per condition and session.

TABLE 7. Experiment 5: Comparison with the entire valid sample

7a. Response-based measurements

Sample	$\%XCW_{CX\ Illusion}$	$\%XCW_{XC\ Illusion}$	$\%XCW_{CX\ Control}$	$\%XCW_{XC\ Control}$
N=25	54.257 (SE 2.266)	47.298 (SE 2.313)	48.297 (SE 2.087)	50.362 (SE 2.106)
N=23	54.533 (SE 2.397)	46.818 (SE 2.414)	48.495 (SE 2.214)	50.377 (SE 2.243)

7b. Psychometric function-based measurements

Sample	$PSE_{CX\ Illusion}$	$PSE_{XC\ Illusion}$	$PSE_{CX\ Control}$	$PSE_{XC\ Control}$
N=25	-0.473 (SE 0.405)	0.607 (SE 0.507)	0.471 (SE 0.422)	-0.086 (SE 0.383)
N=23	-0.502 (SE 0.430)	0.687 (SE 0.534)	0.433 (SE 0.448)	-0.083 (SE 0.408)

Table 7. Descriptive statistics (sample mean and standard error SE) for response-based and Psychometric function-based measurements per adapting condition and session, when considering the entire sample (N=25) or only those 23 participants with valid blocks from both adapting conditions. Note than, since a participant (ID18) lacks blocks of condition CX and another (ID12) lacks data of condition XC, the number of data points contributing to the descriptive results for the entire sample (N=25) is only 24 for each case, but the extra participant is different for each condition. **7a** summarizes response-based measurements (%XCW) and **7b** Psychometric-function based measurements.

CHAPTER 8: EXPERIMENT 6: INSIGHTS ON THE MECHANISTIC BASIS OF THE UNIFORMITY ILLUSION IN DENSITY. EVIDENCE FROM SENSORY ADAPTATION

In the previous chapter we explored the mechanistic basis of the Uniformity Illusion (UI) by using adaptation to a V1-based dimension, such as orientation, and ruled out that UI was associated to sensory changes in the primary visual cortex. In the current chapter we apply the same experimental paradigm to texture-density, a dimension that is processed beyond V1 monocular cells and is closely related to one of the hypothetical mechanisms of UI, namely texture processing. We generated adaptation to a pattern suitable for producing UI on density (whereby physically low peripheral density would be experienced as uniform with the high-density central area) and compared its effects with a physically uniform low-density pattern, as well as to uniform high density. In Experiment 6A, an appearance task was employed, analogous to the orientation task in Experiment 5. The response pattern showed an effect of the central high-density area in reports about peripheral density, but further examination revealed the existence of a non-local, time-invariant effect of physical density, which could explain this result without need for any contribution of UI phenomenology. Experiment 6B utilised a perceptual comparison (2AFC) task to reduce the effect of criterion shifts, but results (albeit underpowered due to small sample size) again suggested a non-local adapting effect of the central area that did not correlate with UI phenomenology. In conclusion, a potential effect of UI on density adaptation was not necessary to explain our results, but its existence could not be entirely discarded either. As the effect of density lacked

spatiotemporal specificity, the adaptation-based paradigm was not suitable for studying the basis of UI on this dimension.[2]

1. INTRODUCTION

In the previous chapter we described an experiment on sensory adaptation under the Uniformity Illusion (UI) on orientation. Results from that experiment showed that the tilt after-effect (TAE) was only ever driven by the physical orientation of the adapting peripheral Gabors, even if such orientation was unseen under UI - and replaced by a different illusory orientation in the global percept. In light of these results, we can conclude that the mechanisms that generate the Uniformity Illusion on orientation do not cause alterations in sensory encoding in the primary visual cortex (V1). In other words, UI on orientation does not operate by direct reconstruction of visual information in low-level sensory areas.

What mechanism, then, is responsible for the Uniformity Illusion? The non-local, non-object-based processing of feature information that characterizes the illusion greatly resembles texture perception (S. C. Dakin & Watt, 1997; Julesz, 1981; Thielscher, Kolle, Neumann, Spitzer, & Gron, 2008), a key mechanism for processing of natural scenes, surface discrimination and figure/background segmentation (Zavitz & Baker, 2014). For instance, a study reported that discrimination of textures defined by orientation signals depended on their summary statistics (mean and variability) (S. C. Dakin & Watt, 1997). Thus, it is possible that, in our experiment on oriented Gabor patterns, perceived

[2] Given that this chapter delves into much detail to finally conclude that little can be concluded about UI on density with an adaptation paradigm, the reader may choose to skim it and move forward to the Discussion in Chapter 9.

uniformity was a result of the inability to resolve the summary statistics between central and peripheral areas. Indeed, texture perception has been related to the phenomenon of peripheral crowding (L Parkes et al., 2001): in crowded ensembles, orientation signals would not be lost to the visual system, but rather pooled by compulsory averaging (L Parkes et al., 2001), while, despite being unresolvable by conscious perception, they would remain able to cause tilt adaptation after-effects, proving that orientation crowding occurs in areas beyond the primary visual cortex (V1) (He, Cavanagh, & Intriligator, 1996). All of this echoes our findings about UI on orientation. In this regard, texture processing has been linked to V2 and V3 areas (El-Shamayleh & Movshon, 2011; Okazawa, Tajima, & Komatsu, 2017; C. M. Ziemba, Freeman, Movshon, & Simoncelli, 2016), by combination of V1-like filter responses (Okazawa et al., 2017). Further evidence for this can be found in a perceptual phenomenon which shares some properties with the Uniformity Illusion, namely perceptual filling-in, whereby a visual trait is perceived in an area despite existing physically only in an adjacent region (Komatsu, 2006). While many examples of filling-in (for instance, involving colour spread) are associated with changes in V1 encoding, texture filling-in has been found to alter encoding only in higher areas, V2 and V3 (De Weerd, Gattass, Desimone, & Ungerleider, 1995; Komatsu, 2006).

Considering all this, we decided to explore the Uniformity Illusion on a visual dimension that is closely linked with texture processing and segmentation of surfaces and objects, namely texture-density (Durgin & Proffitt, 1996; Zavitz & Baker, 2013) - roughly defined as the approximate number of elements per surface area (or spacing between items) Despite its relevance, this visual property is still poorly understood. Adaptation after-effects to texture-density have been consistently reported, suggesting that there may exist neurons responsible for its encoding (Durgin, 1995; Durgin, 1996; Durgin & Hammer, 2001; Hisakata, Nishida, & Johnston, 2016; Sun, Kingdom, & Curtis L. Baker, 2017). Studies on these after-effects have demonstrated that density is a separate property from contrast or spatial frequency (Durgin, 2001; Durgin & Hammer, 2001; Durgin & Huk, 1997) - although it could depend on the combination of outputs from

neurons tuned to high and low spatial frequencies (Steven C. Dakin et al., 2011). On the other hand, there is controversy regarding whether (Steven C. Dakin et al., 2011; Durgin, 2008; M. J. Morgan, Raphael, Tibber, & Dakin, 2014; Raphael & Morgan, 2016; Sun et al., 2017; Tibber et al., 2012) or not (Anobile, Castaldi, Turi, Tinelli, & Burr, 2016; Anobile, Cicchini, & Burr, 2013; Anobile, Cicchini, & Burr, 2016; Anobile, Turi, Cicchini, & Burr, 2015; Arrighi, Togoli, & Burr, 2014; David Burr & Ross, 2008) density and numerosity are based on a common visual metric, with some authors proposing that a change from a numerosity to a texture-density regime depends on the ability to resolve individual elements, which is in turn limited by crowing-like effects and scales with visual eccentricity (Anobile et al., 2015). Furthermore, while early studies on density-related adaptation reported that after-effects only ever decreased perceived density (Durgin, 1995; Durgin & Huk, 1997), suggesting that texture-density was processed as a scalar magnitude -similar to contrast or luminance- (Durgin & Huk, 1997), recent studies on contextual surround effects (simultaneous density contrast) (Sun, Jr., & Kingdom, 2016) and on density after-effects (with test and match presented sequentially) (Sun et al., 2017) have found bidirectional adaptation effects, consistent with a framework of density perception based on channels with sensitivities selective to different values of density. A study has proposed that adaptation to texture-density (spacing between elements of an array) may tune an adaptable metric of the visual space to the properties of the environment, thus affecting subsequent perception of sizes and distances (Hisakata et al., 2016).

As for the brain area(s) responsible for density processing, after-effects show complete interocular transfer, indicating that density coding takes place beyond V1 monocular cells (Durgin, 2001). Although the extent of commonality between density and numerosity is not clear, the loci of numerosity processing might provide some insight: several studies have pointed to a major role for the parietal cortex, specially the intraparietal sulcus (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Zorzi, Di Bono, & Fias, 2011), with an early processing stage possibly in areas V2-V3 (Michele Fornaciai, Brannon, Woldorff, & Park, 2017).

At any rate, the purpose of this experiment is not to solve the questions concerning the properties of density processing, but to explore whether the phenomenological experience of UI on density correlates with density adaptation, in a manner suggestive of texture-density recoding in peripheral fields of visual areas beyond V1.

2. METHODS

Experiment 6 comprised two versions: 6A and 6B. Both of them explored adaptation-induced differences in reports about density patches and whether these differences correlated with experience of UI during adaptation. The main difference between 6A and 6B was the required report: an appearance task in 6A and a perceptual comparison task (2AFC) in 6B. Similar to Experiment 5 (UI on orientation), where participants had to compare the perceived appearance of a Gabor tilt to an internal reference (verticality), in Experiment 6A they had to compare the perceived density of the patch to an internalized reference density that had been previously learnt through training. Because this reference density was arguably more prone to memory distortion than verticality, Experiment 6B supplied an external reference, by asking participants to compare between two simultaneously presented patches -one of them the reference density, the other the test density to be judged.

Henceforth we will describe the methodology of Experiment 6 as a whole, specifying the particularities of 6A and 6B when pertinent.

The structure of Experiment 6 shared multiple similarities with Experiment 5 as described in the previous chapter, including the division into an Illusion session and a Control session, and the division of each experimental block into an adaptation phase

and a test phase - see Figure 23. There were however several differences. In Experiment 6A, before the eight experimental blocks of each session, a training block was run to enable learning and internalization of the reference density (see below). After this training, a practice block helped participants to familiarise themselves with the main task - both in 6A and 6B.

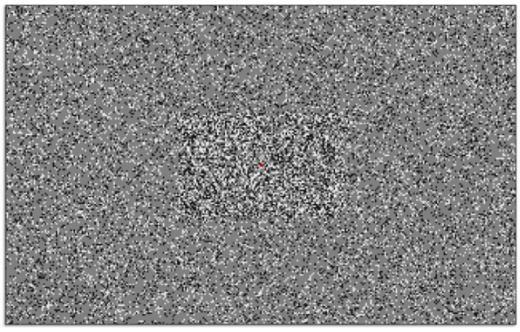
The basic structure of each experimental block was as follows: during the adaptation phase, a dotted pattern was presented on the screen for 120 seconds. The central and peripheral area of the pattern had each a certain density of dots. Participants had to report whenever they perceived the density of the entire pattern as uniform. During the test phase, a perceptual task required comparing the density of a small dotted patch presented in a specific peripheral position with a reference density that was the same for the entire experiment. As mentioned above, in 6A, the reference was internalized by training before the experimental blocks, and a single peripheral patch was presented per trial, requiring participants to report whether its density was lower or higher than the reference. In 6B, two peripheral patches were presented per trial, at the same eccentricity in the left and right hemifield, one of them containing the reference density and the other the trial's test density: participants had to report which of them was higher. Reports were analysed to ascertain adaptation-induced biases in relation with physical and perceived densities.

The eight experimental blocks were of three types, randomly interleaved, differing in terms of the pattern presented during the adaptation phase: Dd, dd and DD - see below for their description. Six blocks were of Dd type (the only type able to cause UI), while one single block was dd and another DD type, serving as a control.

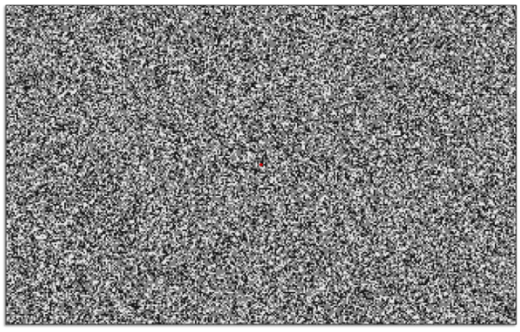
Unless specified otherwise, participant recruitment, apparatus and statistical analyses were the same as in Experiment 5.

ADAPTATION PHASE (120 SECONDS)

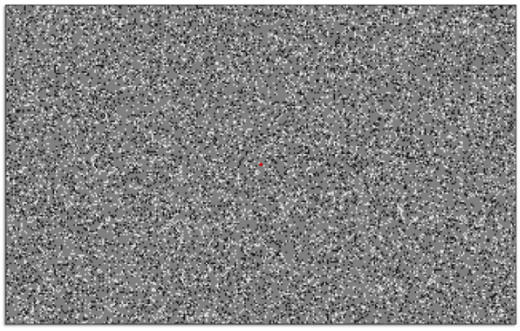
During the **adaptation phase**, in the so-called **Dd blocks** participants are presented with a dotted pattern with high dot density in the centre and low density in the periphery (**Dd pattern**).

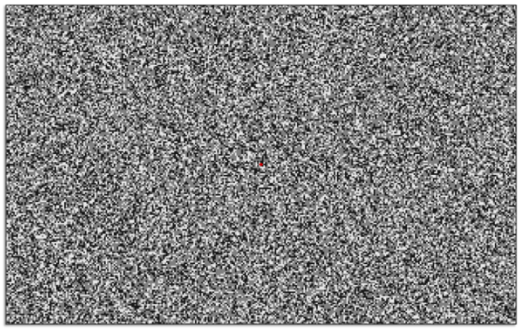


At times, due to the **uniformity illusion** (UI), the entire pattern appears to present high dot density. This **perceptual uniformity** is replaced by a **physically uniform pattern** with high density in centre and periphery (**DD pattern**) during the Control session.



Apart from Dd blocks able to produce UI, each experimental session also has **dd blocks** and **DD blocks**, wherein the adapting pattern is a physically uniform pattern with low (dd) or high (DD) density in both centre and periphery.





TEST PHASE (24 TRIALS, 0.5 SECONDS PER STIMULUS)

Experiment 6A: internal reference – appearance task. During the **test phase**, a small dotted circular region is presented in a peripheral location; participants have to report if its density is lower or higher than an internalized **reference density**, which has been learnt previously by training at the beginning of the experiment. Reports are analysed to ascertain the adaptation effect exerted by physical and illusory density values.

Experiment 6B: external reference – perceptual comparison task (2AFC). During the test phase, two peripheral patches are presented in both hemifields, one of them containing the reference density. Participants have to report which patch has higher density. Reports are analysed to ascertain the adaptation effect exerted by physical and illusory density values.

Figure 23. Experiment 6: Methods. During the adaptation phase, a dotted pattern with certain densities in the central and peripheral regions was presented for 120 seconds. The experiment had three block types, depending on the presented pattern: Dd (high central density, low peripheral density), dd (uniformly low) and DD (uniformly high). Dd pattern was suitable for producing the uniformity illusion (UI), whereby participants saw a homogenous high-density pattern across the entire screen, although physical density in the periphery remained low. This illusory percept alternated with a non-illusory, non-uniform percept at different times during adaptation. The Control session replicated the phenomenology of the Illusion session, replacing perceived with physical uniformity at times in which the participant reported UI in the Illusion session. Furthermore, dd and DD blocks were interleaved with Dd type, to ascertain the effect of adaptation to constant, physically low (d) or high (D) density. The test phase had 24 trials, and required a comparison between a reference density (the same for the entire experiment) and a test density, different for each trial. Both test and reference density were intermediate between the low (d) and high (D) densities employed in the adaptation phase. In Experiment 6A the reference density was internalized by training at the beginning of the experiment, and a single small peripheral patch was presented per trial, containing the test density; participants had to report if it was higher or lower than the reference density. In Experiment 6B the reference and test density were presented simultaneously in two peripheral patches, and participants had to report which was higher.

2.1. PROCEDURE

2.1.1. Training block (Experiment 6A only)

In Experiment 6A, the training block was designed so that participants could internalize a specific density value (reference density), to be used as an internal criterion during the experimental blocks. The training took place at the beginning of each session (Illusion and Control), although participants could repeat it anytime between the experimental blocks, at their own discretion.

At the beginning of the training block, a dotted pattern with the reference density was displayed on the whole screen for 10 seconds. The reference density was defined as 8.66 dots/dva^2, corresponding to a 0.3 proportion (30%) of the surface area covered by dots. After the 10-second exposure, the pattern was removed and two small dotted patches were presented, each on one side of the screen: one of them had the reference density and the other had one of the following values, in terms of proportion of dotted area:

0.225, 0.25, 0.275, 0.325, 0.35 and 0.375 (between 6.50 and 10.83 dots/dva^2). Participants had to report which of the two patches contained the previously displayed reference density. There were 24 trials (4 repetitions of each test density), and the reference was placed 50% of times on each side of the screen. Participants had to repeat the training block if they made more than 10 wrong answers.

2.1.2. Experimental blocks

Illusion session

At the beginning of each block, a pattern formed by black and white dots on a grey background was displayed on the screen. Participants were instructed to report whenever they perceived the density of the dotted pattern as homogeneous by holding down a key, and releasing when the pattern appeared non-uniform. They had to maintain gaze fixation on a red dot located on the screen centre while using an eye-tracking device with their head position stabilized by a chin and forehead rest. If their gaze deviated from fixation by over 1 degree of visual angle (dva), the pattern was removed from the screen and only a grey background showed. The adaptation phase lasted 120 seconds, but actual pattern exposure could be shorter due to gaze-contingent pattern removal. This duration was shorter than for Experiment 5 (UI on orientation - 180 seconds) because during the pilot phase of the current experiment the illusion (described as the experience of a uniform high-density pattern) was reportedly easier to see.

As mentioned above, there were three types of adapting patterns, and thus three types of blocks: Dd pattern had high density in the central area (11.55 dots/dva^2, proportion of covered surface: 0.4) and low density in the periphery (5.77 dots/dva^2, 0.2 proportion of dotted area), and was the only physically non-uniform pattern, suitable for producing UI. Six of the eight blocks in each session were Dd blocks. The other two were each of

one of the remaining types, with uniform density across the entire pattern: dd (uniformly low-density pattern, i.e. 0.2 of dotted area) and DD (uniformly high density, 0.4). These were included as control, to measure the bias exerted by constant exposure to physically high and low density, in contrast with the physically mixed-density Dd pattern (but locally low in the periphery) and the illusory high-density DD pattern seen under UI in Dd blocks.

The test phase had 24 trials, separated by a pseudo-random interval between 750 – 1250 milliseconds. Each trial required a comparison between the reference density, constant for the whole experiment (0.3) and a test density which varied on a trial-by-trial basis. In Experiment 6A, test density could take one of six equally frequent values: 0.225, 0.25, 0.275, 0.325, 0.35 and 0.375 proportion of surface covered by dots (between 6.50 and 10.83 dots/dva^2). In Experiment 6B, the possible values were: 0.215, 0.225, 0.25, 0.35, 0.375, 0.385. Therefore, all values were intermediate between the low (d=0.2) and high (D=0.4) adapting densities, and 50% of them were lower/higher than the reference density (0.3). In Experiment 6A, a small peripheral patch containing the test density was presented for 500 ms on an otherwise empty grey background (except for a red fixation dot on the screen centre). In Experiment 6B, in addition to the test density, an additional peripheral patch containing the reference density was presented for the same time in the opposite hemifield, at the same eccentricity. The test density was presented 50% of trials on each side of the screen.

Control session

Each block of the Control session was designed to be have identical to a homologous block of the Illusion session in terms of phenomenology. During the adaptation phase, the presentation pattern was shown at exactly the same times as in the homologous block: that, is, if the pattern had been removed in the Illusion session due to incorrect gaze fixation, it was removed for the same times in the Control session; besides, it was

also removed if fixation lapsed at other times during the Control session. In dd and DD blocks, the presented pattern was identical at all times; in Dd blocks, whenever participants had reported UI during the Illusion session, the Dd pattern was replaced by a physically uniform, high-density DD pattern at those specific times.

The test phase was identical to the homologous block in the Illusion session. The inter-trial interval was adjusted in function of the response time in order to make the intervals between successive presentations equal in the Illusion and Control sessions.

2.2. STIMULI

Stimuli were displayed on a grey background. A red fixation dot (0.42 dva diameter) showed constantly on the screen centre.

The visual dimension of density was built by assembling a collection of black and white dots - 50% of each, to ensure isoluminance between areas, patterns, density values and with the grey background (~34.8 cd/m^2). Each individual dot had a diameter of 0.21 dva and occupied a square cell (0.21 dva side length) on an invisible grid spanning the entire screen. This prevented overlapping between dots. The number of occupied cells would depend on the desired density value, which was defined, as previously mentioned, as the proportion of surface area covered by dots (a 0-1 ranging value), which scaled linearly with to the number of dots by unit of surface, since all the dots had the same size. Once the desired number of dots within a certain area was set (for example, the central area, the test patch, etc), the actual position of the dots was randomized among all the cells in the area.

2.2.1 Adapting pattern

The Dd adapting pattern had a central area, spanning 16 dva horizontally (8 to each side) and 12 dva vertically, wherein dot density was 0.4 (11.55 dots/dva^2) - high density, D. The rest of the screen comprised the peripheral area, with 0.2 dot density (5.77 dots/dva^2) -low density, d. In the dd pattern, density in both central and peripheral areas was 0.2; in DD pattern it was uniformly high, 0.4.

The position and colour of the individual dots did not rest fixed throughout the entire adaptation period, but were reshuffled at periods randomized between 180 – 300 ms (18 – 30 frames). Only the overall number of dots in the area (which determined the dot density) remained the same.

As described above, in Dd blocks of the Control session, the Dd pattern was replaced by a uniformly high DD pattern at the times wherein UI was reported in the homologous Illusion block.

2.2.2. Test patches

During the trial phase, a circular patch of 3 dva diameter was presented at 13.5 dva from the screen centre, either on the left or the right hemifield along the horizontal axis - therefore, the minimum distance between the inner region of the patch and the external border of the central area was 4 dva (=13.5-1.5-8). This circular area was covered by black and white dots up to a given test density, always intermediate between 0.2 and 0.4: specifically, 0.225, 0.25, 0.275, 0.325, 0.35 and 0.375 in Experiment 6A, or 0.215, 0.225, 0.25, 0.35, 0.375, 0.385 in Experiment 6B. In Experiment 6B, a second circular patch was presented in the opposite hemifield, also at 13.5 dva eccentricity, with the reference density (0.3).

2.3. STATISTICAL ANALYSIS

Methods for statistical analysis were identical to Experiment 5 (UI on orientation) except for the specific details of Psychometric curve fitting: in Experiment 6A, a cumulative Gaussian curve was fitted by the method of maximum likelihood to the proportion of 'higher' reports (higher than reference density) per value of test density, separately for each participant, session and adapting condition (dd, Dd, DD). The threshold (α) for 0.5 proportion of 'higher' responses and the slope of the curve (β) were free parameters (starting values: α=0.3 (30% density), β=0.5), while guessing (γ) and lapse rate (λ) were fixed at zero. In Experiment 6B, the stimulus intensities for the Psychometric curve were the absolute difference between test and reference density, and the corresponding value of the function was the proportion of correct reports (when choosing the higher density between test and reference). The threshold and slope were free parameters (starting values: α=0.075, β=1), while guessing (γ) and lapse rate (λ) were fixed at 0.5 and 0, respectively. Nevertheless, Psychometric curves were not key to our analyses (especially in Experiment 6B) as the data was difficult to fit due to the narrow range of test densities imposed by experimental constraints.

3. RESULTS

3.1. EXPERIMENT 6A: APPEARANCE TASK

Twenty participants volunteered to the experiment (14 female, mean age 24.6 y/o).

As in Experiment 5, we sought to ensure sufficient exposure to the adapting pattern by excluding blocks where it had been presented for less than 2/3 of the adaptation phase (<80 seconds), as well as the homologous blocks to those excluded for the reason above.

This led to exclusion of 70 blocks out of 360 (21.88%), including the entire datasets from 3 participants; results presented here correspond to the included blocks of the remaining 17 participants.

3.1.1. Adaptation phase

Data reported here refer to Dd blocks only, with an adapting pattern suitable for producing UI.

On average, time of effective exposure to the adapting pattern per block was 109.08 and 97.77 seconds during the Illusion and Control session, respectively, making 90.90% and 81.47% of the total time of the adaptation phase (120 seconds).

Perceived uniformity was reported, on average, for 59.11 seconds during the Illusion session and 50.74 seconds during the Control session, making 53.84% (minimum 16.38%, maximum 90.92%) and 51.54% (minimum 17.75%, maximum 89.48%), respectively, of the total time of pattern exposure. These were far greater proportions than those for UI on orientation in Experiment 5 (~27% on average, with one participant as low as ~0.55%). The average proportion of reported uniformity was not significantly different across sessions, with anecdotal evidence in favour of equality of proportions: paired-samples t-test, BF_{01}=2.122. Physical uniformity during the Control session was reported as perceptually uniform 72.42% of the time, while physical non-uniformity was identified as uniform only 17.28% of the time, compared to 53.84% during the Illusion session, when the presented pattern was also physically non-uniform. As in Experiment 5 with orientation patterns, it is likely that presentation of a truly uniform pattern at times during the Control session produced a conservative shift in participants' criterion for reporting uniformity. This highly suggests that decisional biases play a role in the phenomenology of UI.

3.1.2. Hypotheses and measurements

As described in the Introduction to the current chapter, the mechanisms involved in processing of texture-density are less well characterized than orientation processing. For this motive we defined three competing hypotheses *a priori*:

1. Physical and local adaptation. Adaptation to the density of the dotted pattern is driven solely by physical density *and* is local (retinotopic) in scope. Under this scenario, adaptation-induced response bias caused by exposure to Dd pattern would be the same as for dd pattern, since both patterns have low density in the peripheral area where test patches are later presented. In both cases, there would be a relative overestimation of test densities compared to DD blocks. This would be the case even when uniform high density ('*illusory DD'*) is perceived for very long times under UI, in Dd blocks.

2. Physical, non-local adaptation. Adaptation to the density of the dotted pattern is driven solely by physical density, but is non-local in scope, either because it has very large receptive fields or it is processed as an abstract dimension. Under this scenario, the high-density central area would exert *some* bias on reports about the density of the peripheral patches presented later on during the test phase, but this effect would be due to the physical presence of this central area and unrelated to perception of UI. Under this hypothesis, response patterns for dd and Dd blocks should be different, with dd showing overestimation compared to Dd. The most likely prediction is that both central and peripheral areas would have a measurable effect, and therefore the response pattern in Dd blocks would be intermediate between dd (overestimation) and DD (underestimation of test density). At any rate, since the perceived density under UI has no effect, we would expect independence of time of perceived uniformity.

3. Illusion-driven adaptation. The perceived uniform high density for the Dd pattern under UI ('Illusory DD') is able to cause adaptation effects by recodification of density information in peripheral receptive fields. Thus, we should expect a similar response pattern than for the previous hypothesis (non-local physical adaptation) when considering overall results, but in addition there should time-dependency of the duration of UI. The most likely prediction for Dd blocks would be an intermediate response pattern between dd (overestimation) and DD (underestimation), with a trend toward greater underestimation for longer times of UI.

As in the orientation experiment, we based our analyses on direct responses as well as fitted curves:

1. The proportion of 'higher' reports (test density is deemed higher than the reference density) was termed %Higher. Since 50% of presented test densities are below and above the reference density, an unbiased observer would report 'higher' 50% of times. Thus, %Higher>50% indicates response overestimation and vice versa.

2. The point of subjective equality, PSE, was defined as the test density that elicited 50% 'higher' (than reference density) reports, and as many 'lower' reports. Since the reference density is 0.3 (proportion of surface covered by dots), a PSE<0.3 indicates overestimation and vice versa.

As described above, we may make different predictions under the three different hypotheses:

1. Physical and local adaptation: $\%Higher_{dd}=\%Higher_{Dd}>\%Higher_{DD}$, $PSE_{dd}=PSE_{Dd}<PSE_{DD}$. Results are independent of time of perceived uniformity.

2. Physical, non-local adaptation: $\%Higher_{dd} > \%Higher_{Dd} > \%Higher_{DD}$, $PSE_{dd} < PSE_{Dd} < PSE_{DD}$. Results are independent of time of perceived uniformity.

3. Illusion-driven adaptation: if UI is perceived for less than 100% of the time we should expect $\%Higher_{dd} > \%Higher_{Dd} > \%Higher_{DD}$, $PSE_{dd} < PSE_{Dd} < PSE_{DD}$. Results are *dependent* of time of perceived uniformity.

In the following sections we will present both response-based and Psychometric function-based results but our analyses will focus on the first to avoid redundancy, as the same pattern of results is observed for both approaches.

3.1.3. Responses show an effect of the central high-density pattern, but are independent of the reported time of UI

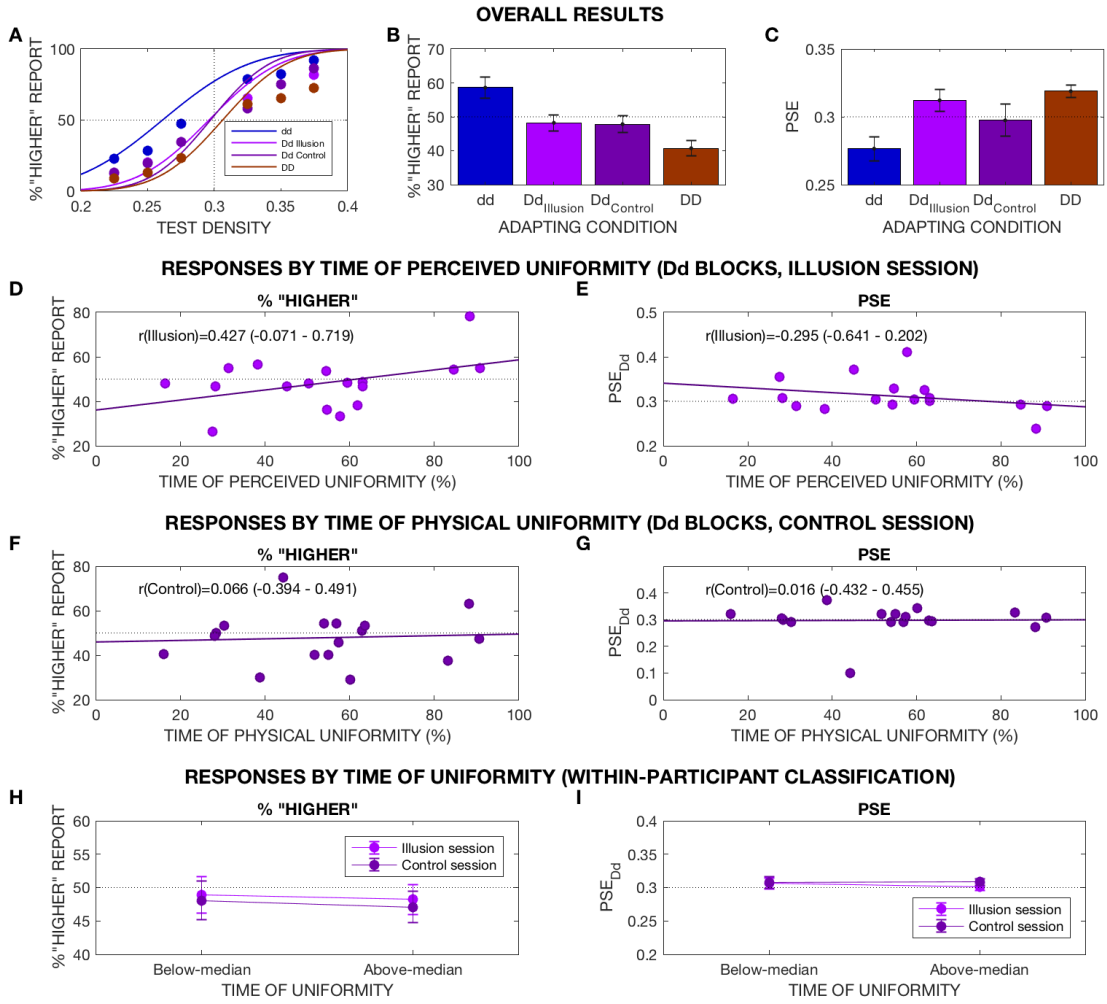


Figure 24. Experiment 6A: Results. Reports about density are influenced by both the local low density *and* the central high density present in Dd pattern during the adaptation phase. However, time-invariance from both perceived and physical uniformity prevents resolving whether this result is a non-local effect of the central pattern or an effect of the illusory high density perceived during UI. **24a – c:** Overall results. Reports about density in Dd blocks are virtually unbiased, while dd and DD blocks show a noticeable over and underestimation, respectively, consistent with adaptation effects - or a criterion shift as a result of exposure to the pattern. This result rules out that the effect of exposure to density is both physical and local-driven, since in that case we would expect identical response patterns in Dd and dd blocks (low density in the peripheral area). Thus, two possibilities remain: physical, non-local effect or illusion-driven effect. Results in the Control session seem to support the latter, since they show almost identical

response pattern to the Illusion session despite presentation of physical uniformity at times. But, if a potential non-local effect of density was also largely time-invariant, the similar results in Illusion Dd and Control Dd blocks would be explained without need for a specific effect of UI. In this regard, the analysis of time-dependency is crucial. **24d – 24e:** Time-dependent results during the Illusion session. The two graphs present bivariate correlations between each participant's time of perceived uniformity (as a percent of the total time of pattern exposure) and their response bias in Illusion Dd blocks, expressed in terms of proportion of 'higher' reports (**24d**) and PSE (**24e**). Thus, each data point represents a single participant. If UI had an effect on responses and this effect was time-dependent, we should observe a trend to greater underestimation for longer times of perceived uniformity, i.e. a negative correlation with %Higher and a positive correlation with PSE. Analyses entirely discard this, rather showing a non-significant trend in the opposite direction. **24f - 24g:** Time-dependent results during the Control session: bivariate correlations between individual time (%) of physical uniformity and response bias in Control Dd blocks. As for the Illusion session, time-dependency would produce a negative correlation between time of uniformity and %Higher, and a positive correlation with PSE. Again, results show time-invariance of response pattern in Dd blocks, even in relation to physical replacement at times by a uniform high-density (DD) pattern. **24h-24i:** Time-dependent results, analysed within-participant: each participant's dataset has been divided in terms of whether each Dd block was below or above the participant's median time of (perceived/physical) uniformity. The graphs show that the participant's response pattern (expressed in terms of % 'Higher' reports or of PSE) is not different in blocks with short compared to long time of uniformity, neither for perceived nor for physical uniformity. Crucially, the fact that the effect of exposure to physical density is also time-invariant (or not sensitive enough for our analyses), as shown in the Control session, means that our experiment cannot disambiguate between the possibilities of non-local, physical (and time-invariant) adaptation and UI-driven (time-invariant) adaptation.

Overall results

Figure 24a presents the sample's average proportion of 'higher' (than reference density) reports for each test density value, plotted separately by adapting condition: dd, Dd (Illusion session), Dd (Control session) and DD. Block types dd and DD are pooled across sessions, since there are no differences between Illusion and Control. In Dd blocks during the Control session, the Dd pattern is replaced by a physical DD pattern during the times of reported UI in the Illusion session.

For illustration-of-method purposes, Figure 24a presents cumulative Gaussian curves fitted for the sample's pooled data (N=17). However, data analyses are based on

measures obtained individually on each participant: %Higher and PSE, summarized in Figures 24b-c.

Illusion session

We will begin by comparing the adapting conditions dd, Dd (Illusion) and DD, setting momentarily aside Dd blocks during the Control session, which presented a mixture of Dd and DD patterns at different times.

Figures 24a-c show that responses were influenced by the pattern presented during the adaptation phase: there was a noticeable overestimation of test densities in dd blocks (%Higher=58.60%), virtually no bias in Dd blocks (%Higher=48.23%) and underestimation in DD blocks (%Higher=40.74%). A Bayesian RM ANOVA on %Higher with adapting pattern (dd, Dd Illusion, DD) as within-participant factor showed extreme evidence in favour of the explanatory role of the adapting pattern (BF_{10}=118.298). PSEs show an analogous pattern of response bias (Figure 24c): PSE_{dd}=27.65, PSE_{Dd}=31.22, PSE_{DD}=31.89.

If density adaptation was physical *and* local-driven, the high-density central area in Dd pattern should not have any effect on responses, since the test patches are presented in the periphery. Conversely, if the central high-density area has some kind of effect (regardless of whether it is physical or illusion-driven), we should expect a relative underestimation of test densities in Dd pattern, compared to dd: $\%Higher_{dd} > \%Higher_{Dd}$. A Bayesian paired samples t-test rendered moderate evidence in favour of this prediction: BF_{+0}=3.388. In short, evidence indicates that the effect of pattern exposure is not simultaneously physical-driven *and* local in scope: rather, the high-density central pattern in Dd has an influence on partipants' responses. Whether it is a physical-driven, but non-local effect, or an illusion-driven bias, is still unclear.

Control session

During the adaptation phase, Dd blocks in the Control session present an alternation of Dd and DD patterns, depending on the phenomenology reported during the Illusion session.

The proportion of 'higher' reports in Control Dd blocks is 47.84%; statistically, it is no different than for Illusion Dd blocks (48.23%): Bayesian paired-samples t-test BF_{01}=3.997 (moderate evidence for the null hypothesis). PSE-based analyses display the same pattern: $PSE_{Dd\ Illusion}$=31.22, $PSE_{Dd\ Control}$=29.74, Bayesian paired-samples t-test with moderate evidence in favour of equality (BF_{01}=3.108).

Thus, contrary to the Experiment 5 (UI on orientation), here results suggest that the physical replacement of illusory by physical uniformity (illusory DD by physical DD pattern) accurately replicates the adaptation-driven response bias observed for a pattern suitable for producing UI on density. This may be deemed supportive of an effect of UI on adaptation to density, but, alternatively, if the non-local effect of density was also time-insensitive, this pattern of results would not need to assume any effect of UI, as both Illusion and Control Dd blocks present a mixture of low and high densities, differing only in time and physical location of each.

Time dependent results

Results so far clearly rule out that the response bias driven by exposure to Dd pattern is solely based on the local effect of the low density physically presented in the periphery. However, two possible interpretations of our data remain: (i) either the high-density central area exerts some non-local, but physical-driven effect (i.e. the physical presence of a high-density pattern biases participants' responses for a patch in a different retinal location, due to very large receptive fields, non-local adaptation effects or shifts in

decision criteria), or else (ii) the perceived high density in the periphery under UI (illusory DD pattern) causes a bias in responses.

Results from the Control session point to the latter possibility, since replacement of illusory DD percept by physical DD pattern in function of participants' illusory experience renders similar results to the Illusion session. However, finding time-dependency during the Control session should be a critical requirement for determining our ability to disambiguate between both hypotheses (non-local physical versus UI-driven effect) with the current experimental design. If there is such time-dependency in the Control session, then whether or not there is time-dependency *also* in the Illusion session would resolve the alternative between the remaining hypotheses: in this scenario, finding time-dependency would indicate a UI-driven effect; conversely, time-*in*dependence of illusory uniformity (in presence of time-dependency of physical uniformity) would mean that the Dd-driven response bias is due to a non-local effect of the physical high-density area.

However, if we failed to find time-dependency of physical uniformity, it would mean that perceptual decision-making on density is affected by contextual density in a manner that lacks the needed spatiotemporal sensitivity to provide informative answers about the basis of UI. Indeed, UI needs the physical presence of both high and low-density areas, and it is not perceived 100% of the time; therefore, non-local, time-invariant (or time-insensitive) effects of the physical pattern would resemble a hypothetical UI-driven effect in a way that could not be easily disambiguated in our data - both cases would entail an intermediate response bias between low and high-density effect. In that case, the overall results of the Control session (which seemed to favour a UI-driven effect on responses) could be explained just in light of such invariance.

We therefore proceeded to analyse time-dependency in a similar way than for Experiment 5, i.e. by examining the bivariate correlation between each participant's time of perceived/physical uniformity and their individual response bias. This analysis

focused on Dd blocks only, as they were the only ones wherein UI could be perceived. The existence of time-dependency in relation to perceived and/or physical uniformity (illusory or physical DD pattern) would predict a trend toward increased underestimation of test density for a longer time of uniformity: in other words, a negative correlation with %Higher and a positive correlation with PSE_{Dd}. Conversely, time-invariance would be expressed by a lack of any statistical association whatsoever.

Results for the Illusion session show a trend toward a *positive* correlation between time of perceived uniformity and proportion of 'Higher' reports, contrary to what would be predicted by an adaptation-like effect of the illusory percept; however, evidence for the existence of this correlation was only anecdotal: r=0.427, 95% credible intervals: (-0.071, 0.719), BF_{10}=1.157. If we used a one-sided negative prior according to our prediction of a negative correlation, we observed moderate evidence against such prediction: BF_{-0}=0.123. Converging results were obtained when using PSE_{Dd} as a measure of response bias. In short, we observed no trace of time-dependency of illusory uniformity in the manner predicted by UI-driven adaptation effect. See Figures 24d and 24e.

Concerning the Control session, a Bayesian Pearson's correlation between time and %Higher provided moderate evidence against the existence of any statistical association: r=0.066 (-0.394, 0.491), BF_{10}=0.309 - see Figure 24f. Evidence against a negative correlation (what would be predicted by time-dependent adaptation to physical high density) was also moderate: BF_{-0}=0.250. PSE-based correlation again showed converging results (Figure 24g). Thus, the effect of exposure to different physical densities was time-invariant, or at least not time-sensitive enough in the conditions of our experiment.

We sought to further confirm this lack of association by analysing within-participant differences between experimental blocks of the same session. With this we aimed to remove the noise caused by participants' idiosyncratic biases present in the between-participant correlations. We divided the Dd blocks of each participant and session into

two parts, according to whether time of (perceived/physical) uniformity had been below or above the participant's own median time, and ran a Bayesian paired-samples t-test for the difference in the proportion of 'higher' reports in below-median compared to above-median time blocks. Results are shown in Figures 24h-24i: again, evidence supports equality of response proportions in both cases, for both the Illusion (BF_{10}=0.321) and the Control session (BF_{10}=0.261).

In conclusion, density reports in our experiment are independent from the time of perceived and physical uniformity. Time-invariance from exposure to both illusory *and* physical density (as seen in the Control session) means that density processing in our experimental design lacks the necessary retinal specificity and temporal sensitivity to be employed as a tool for exploring the basis of UI through adaptation effects.

3.1.4. Presentation of a density pattern only in the central area demonstrates non-local effects on density reports

Results so far are consistent with two hypotheses: physical, but non-local and time-invariant effect of exposure to density, or UI-driven effect (also time-invariant). The lack of time dependency prevents deciding between both possibilities. However, it is possible to ascertain the existence of non-local effects in a pattern unsuitable for producing UI. If we found a similar response bias is such conditions, we would have to conclude that non-locality is sufficient to explain our results.

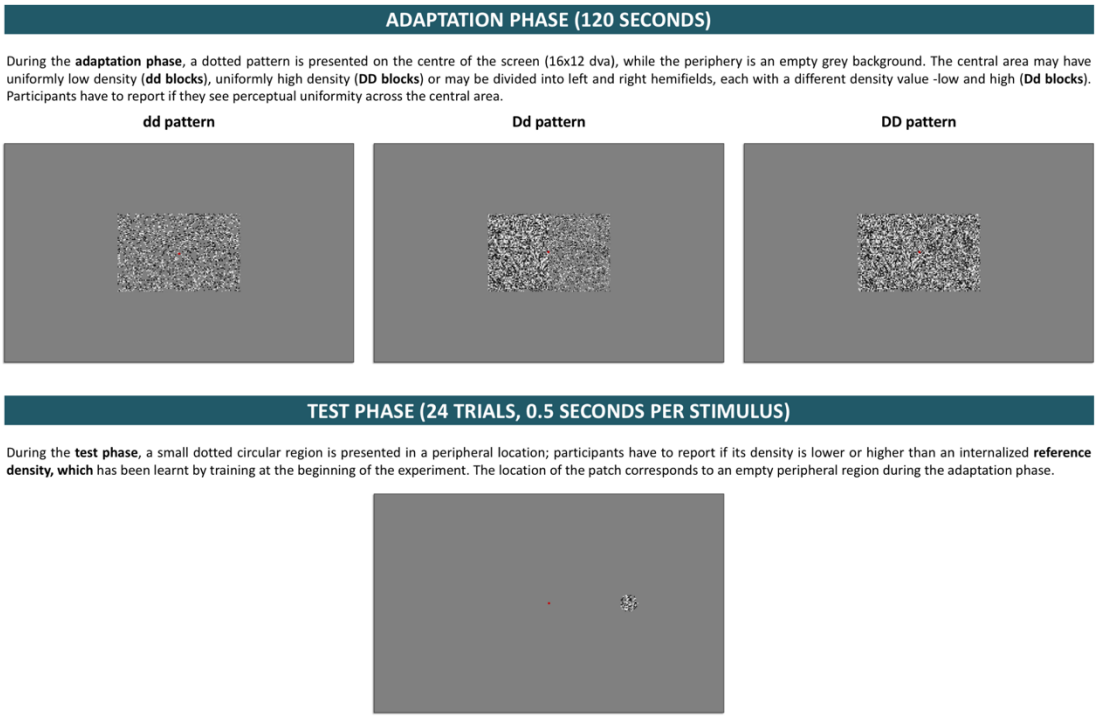


Figure 25. Experiment 6A, Central session: Methods. During the adaptation phase, a dotted pattern is presented on the central area of the screen; according to its density, the block would be termed dd (low), DD (high) or Dd (split into two halves by a vertical axis, with low or high density in each). The peripheral region was always an empty isoluminant grey background. The test phase was identical to the Illusion and Control sessions in Experiment 6A, requiring a comparison of a presented peripheral density with a reference density previously internalized.

We therefore designed an additional experimental session, termed 'Central' session, wherein the adapting pattern was circumscribed to the central area (a 16x12-dva rectangle centred on the screen). The peripheral area only displayed a grey, isoluminant background. In Central Dd blocks, the left and right half of the central area presented different dot densities: low (0.2, 20% of dotted area) and high (0.4) density. Which density was presented in each side was randomized per block. In Central dd blocks, a uniform low density was displayed in the central area; in Central DD blocks, it was a high-density pattern. Thus, this design presented uniformly low, uniformly high and a mixture of low and high densities, same as the corresponding blocks in the Illusion session, but

presentation here was always non-local, since the peripheral area, where the test patches would be displayed during the test phase, was empty. The Central Dd pattern was also unsuitable for producing UI, although other similar processes (such as perceptual filling-in, etc) might potentially take place. Other than the adapting pattern, the experimental design of the Central session was identical to the Illusion session (reported above). Participants were given identical instructions in both parts, albeit specifying that uniformity in the Central session only needed to occur within the central area.

Nine participants volunteered to this experiment (seven female, mean age 23.2); five of them were participants who had done the Illusion and Control sessions of the main experiment and volunteered to an extra Central session. The other four did a two-session experiment, with an Illusion and a Central session instead of a Control session.

In Central Dd blocks, participants reported perceptual uniformity across the central area during 14.52% of the time (13.55 s), compared to 56.48% for Illusion Dd blocks.

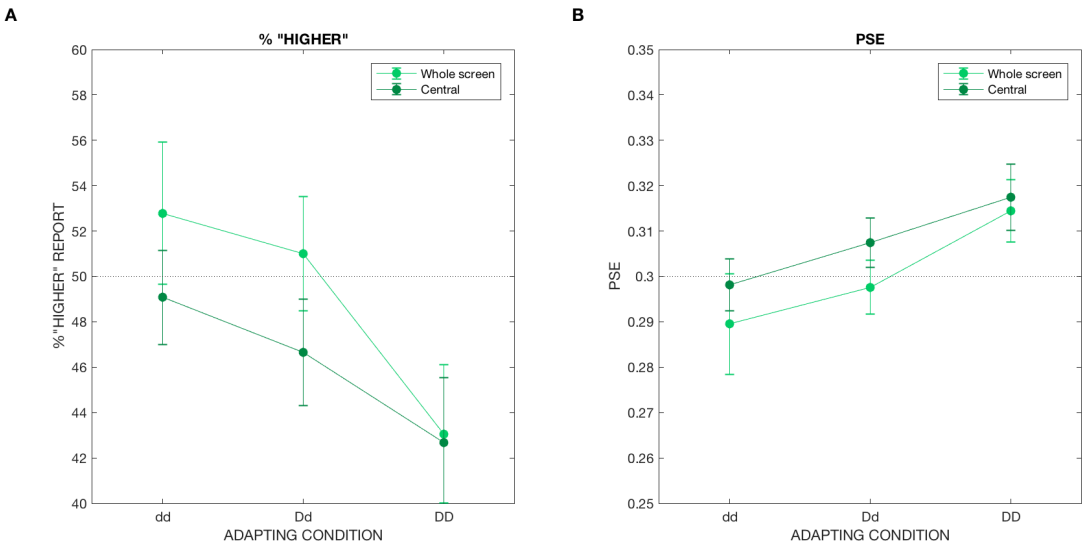


Figure 26. Experiment 6A, Central session: Results. Response bias in function of adapting pattern (dd, Dd, DD) and scope of the display (whole screen, central). Figure 26a presents response-based measures of bias, whereas 26b shows PSEs. By both approaches we observe a non-local effect of the densities presented in the central area, similar to that observed for the whole-screen presentation.

Figure 26 shows the pattern of responses in function of the densities presented during the adaptation phase (only low, both low and high, only high density) and the scope of the presentation (the whole screen in the Illusion session, the central area in the Central session). Figure 26a presents results in terms of the proportion of 'higher' reports (%Higher) and Figure 26b in terms of PSEs. Visual inspection of the plots shows that the pattern observed in the Illusion session also appears in the Central session: overestimation in dd blocks, underestimation in DD blocks and an intermediate pattern in Dd blocks - even if it seems slightly less marked in the Central session.

To further confirm this, we ran a Bayesian RM measures ANOVA on proportion of 'higher' reports (%Higher), with two within-participant factors: adapting pattern (dd, Dd, DD) and display scope (whole screen, central). Based on the evidence of the Bayesian ANOVA, a comparison between five possible models was performed: the null model (only containing participant information), two models containing a single main effect (either pattern or scope), a model containing both main effects and the full model with two main effects and the interaction pattern*scope. The key term for our analysis is the interaction term, which determines whether responses are modulated differently by the adapting pattern in function of its presentation in the whole screen or only in the central area. If evidence opposes inclusion of the interaction term in the model, it would indicate that the bias exerted by adapting densities is non-local.

The best model according to evidence of the Bayesian RM ANOVA contained only 'adapting pattern' as predictor (BF_{10}=3.706: moderate advantage over the null). Conversely, the full model containing both main effects and the critical interaction term had a BF_{10}=0.519. Evidence was moderately against inclusion of the interaction term in

the model: $BF_{Inclusion}$=0.289. Thus, the adapting pattern had an explanatory role on participants' response bias regardless of whether this pattern was presented in the whole screen or only the central area. Similar results were obtained when analysing response bias by PSE instead of %Higher.

In conclusion, we have shown that non-local presentation of the adapting densities in a pattern unsuitable for producing UI was sufficient to explain the pattern of density-driven biases observed in Experiment 6A. While results of the Central session do not categorically exclude any possible UI-driven effect during the Illusion session, they make it unlikely as every aspect of our data is well accounted by a non-local, time-insensitive processing of density. They also suggest that adaptation-based experiments on visual dimensions that are not strictly low-level are unsuitable for studying the basis of UI.

3.2. EXPERIMENT 6B (PERCEPTUAL COMPARISON TASK)

Results from Experiment 6A indicate that the different densities presented during the adaptation phase exert a non-local, time-invariant repulsive effect on appearance judgments about density during the test phase. This is consistent with an effect of adaptation by a mid to high-level feature-dimension - i.e. a repulsion of the test density away from the adapting pattern. However, a second possibility is that this effect is caused by a shift in a binary decision criterion for low/high report, due to corruption of the memory representation of the reference density by the density or densities presented during the adaptation phase - an attraction of the *reference* density toward the adapting pattern. To disambiguate between these possibilities, in Experiment 6B we implemented a perceptual comparison (2AFC) task, presenting test and reference density simultaneously and at the same peripheral eccentricity. Such task would be more robust to criterion shifts and would ensure that the effect of the adapting pattern (if any) acted on the same direction upon test and reference densities.

Nine participants were recruited for this experiment (five females, mean age 25.33 y/o), in similar conditions as for Experiment 6A. Results presented here correspond to 101 blocks of 8 participants, after exclusion of all blocks where effective pattern presentation had lasted for less than 2/3 of the adaptation phase (<80 seconds), as well as their homologous Dd blocks in the other session.

A critical caveat: results for this experiment are greatly underpowered due to the small sample size. Recruitment was stopped (due to change in priorities and time constraints), as preliminary results appeared to confirm that the employment of density was unsuitable for investigating the basis of UI, due to its (apparent) non-local and time-invariant effect, also in a perceptual comparison task. For this reason, results remain entirely preliminary.

3.2.1. Adaptation phase

Results pertaining the adaptation phase were very similar to Experiment 6A. Average time of exposure to the adapting pattern in Dd blocks was 109.77 and 97.55 seconds, or 91.48% and 81.30% of the duration of the adaptation phase, respectively for the Illusion and Control session.

Perceived uniformity was reported during 64.71 and 47.99 seconds during the Illusion and Control session, respectively, i.e. 58.78% (27.57% - 97.19%) and 48.59% (12.60% - 95.67%) of the total time of pattern presentation. These proportions were similar to Experiment 6A. On average they were lower for the Control session (although evidence was neither for nor against a difference across sessions: Bayesian paired-samples t-test, BF_{10}=0.929), which, as in previous experiments, may point to a conservative shift in decision criterion for perceptual uniformity motivated by the exposure to true physical uniformity. Physical uniformity during the Control session was reported as uniform

57.53% of the time, while non-uniformity during the same session was for only 16.31%, despite being physically identical to non-uniformity during the Illusion session (reported as uniform 58.78% of time).

3.2.2. Hypotheses and measurements

As mentioned above, the purpose of this experiment was to disambiguate whether results in Experiment 6A were due to adaptation or solely to decision – memory effects.

If the effect found in Experiment 6A was entirely due to a criterion shift for binary labelling we should not see any difference between adapting patterns in the current 2AFC task. Conversely, if there is an effect of adaptation, predictions are less straightforward as many properties of density adaptation are still disputed: whether it is uni or bidirectional, its reference frame, etc. We could make the general prediction that, if there is a non-local effect of the central density beyond criterion shifts, response patterns should be different for dd and Dd blocks; furthermore, if such non-local effect was due to UI, the analysis of time-dependency in the Illusion and Control session should proceed similarly to Experiment 6A.

For our data analysis, we calculated the proportion of correct responses (correctly identifying the patch with the higher density) as a function of the difference between test and reference density. The difference in absolute value (regardless of which is higher) could take three values: 0.05 (5%), 0.075 and 0.085, given the employed test densities (0.215, 0.225, 0.25, 0.35, 0.375, 0.385). Datasets were split according to whether the test density was higher or lower than the reference, as the effect of the different adapting densities was likely not symmetrical. Thus, ‘lower than reference’ comprises test values 0,215, 0.225 and 0.25, and ‘higher than reference’ corresponds to the other three.

3.2.3. Response patterns *suggest* a non-local adaptation effect by the high-density central area, without evidence for time-dependency of perceived or physical uniformity

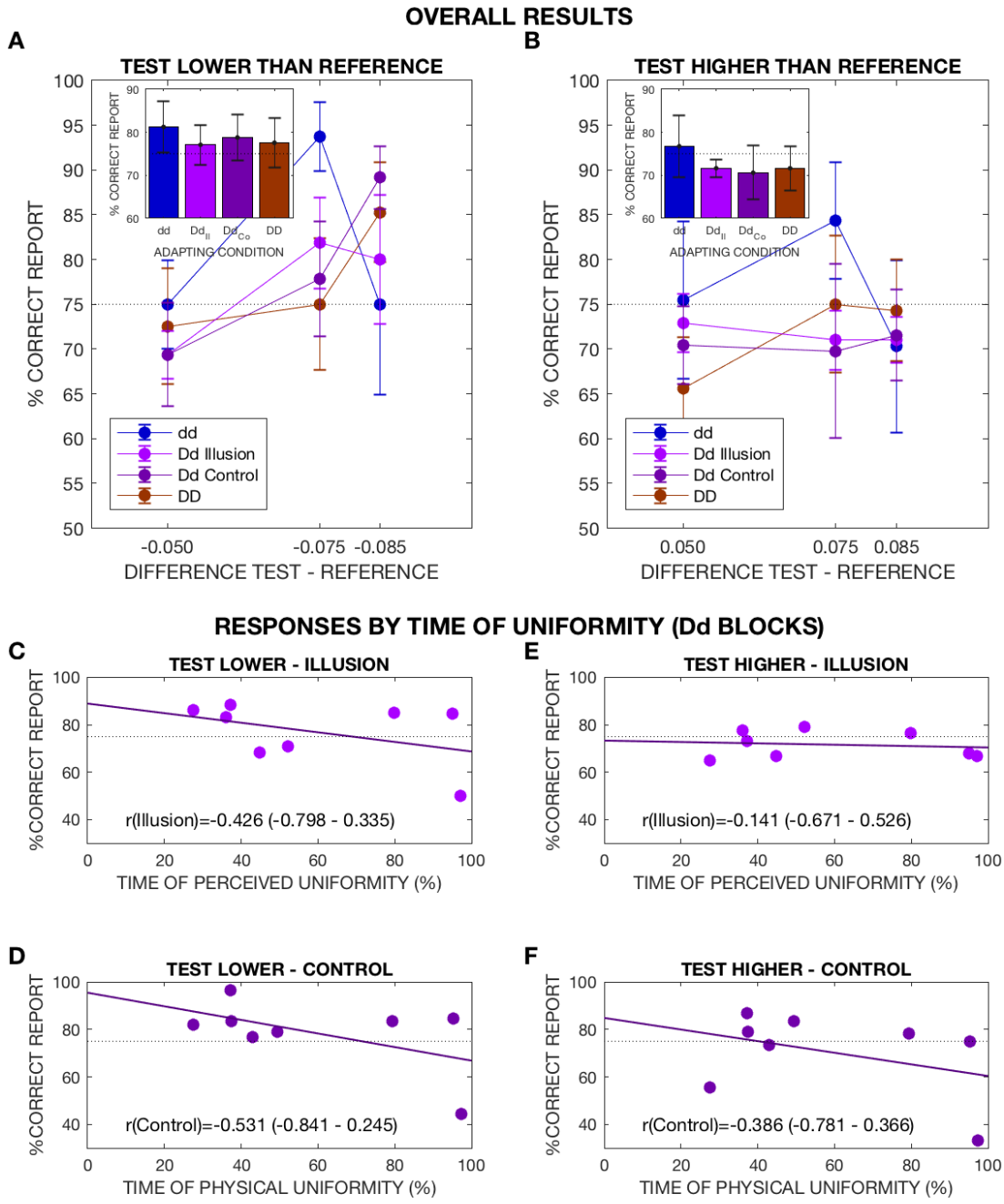


Figure 27. Experiment 6B: Results. Figures **27a** and **27b** present the proportion of correct reports (%) as a function of the stimulus difference between the elements of comparison in the 2AFC task, i.e. test and reference density. Results are split by adapting pattern (dd, Dd - Illusion session, Dd – Control session, DD) and by lower and higher (than reference) test densities. The inset graphs show the total proportion of correct responses per adapting pattern. Results are greatly underpowered, but appear to suggest a non-local, unidirectional adaptation effect of the high-density pattern, as performance is worsened after exposure to any adapting pattern containing high density (even Dd pattern, wherein local adapting density is low in the periphery). This suggests a compression of the distance between test and reference densities toward the lower region of the perceptual space - reducing discrimination and thus performance. Such compression would be exerted by repulsion away from the high density presented during the adaptation phase in Dd and DD blocks – although it is unclear whether the effect of Dd pattern is driven by non-local physical adaptation or by changes induced by UI in the periphery. Conversely, low adapting density (d) would not have such compressive effect as it is itself lower than any of the test or reference densities. Furthermore, the finding of a similar pattern of results for test densities lower and higher than reference suggests that density adaptation is unidirectional. If the adaptation effect was bidirectional (repulsion away from the adaptor, toward lower and higher regions in the perceptual space) we should see a symmetrical effect so that the dd pattern would act on lower test densities in a similar manner as the DD pattern for higher test densities (as our design is symmetrical). Rather, we see the same difference between dd and rest of adapting patterns for lower and higher test densities. **27c – 27f**. Time-dependency analyses: bivariate correlations between time of perceived or physical uniformity and proportion of correct reports in Dd blocks in the Illusion (**27c**, **27e**) and Control (**27d**, **27f**) session, respectively, split by test density values: lower (**27c**, **27d**) and higher (**27e**, **27f**) than the reference density. A non-significant negative correlation appears to confirm the existence of time-dependency for both the Illusion and Control session, which would indicate that the different effect of Dd and dd patterns in the overall results was driven by sensory changes induced in peripheral receptive fields by UI, and not just by non-local effect of the physical high density presented in the central area. However, on closer inspection, time-dependency disappears entirely if we remove a single outlier who consistently performed below chance. Rather, we seem to have (possible, but very underpowered) time-invariance in both the Illusion and Control sessions. As in Experiment 6A, the latter prevents using time-dependency analysis in the Illusion session for investigating the effect of UI.

Figures 27a and 27b show the proportion of correct reports as a function of the difference between the elements of comparison, namely test and reference density, plotted separately by adapting pattern and split according to whether the test density was lower (27a) or higher (27b) than the reference density. The inset graphs present the proportion of correct reports (across all test-reference differences) by adapting pattern, for test densities lower (27a) and higher (27b) than reference density. On visual inspection, in both 27a and 27b, there seems to be a slight benefit of adaptation to dd pattern, entailing a higher proportion of correct reports than any other adapting

pattern, especially for small differences between test and reference density (0.05 and 0.075), i.e. for those cases where the effects of an increased sensitivity would be more noticeable. However, this supposed difference is entirely non-significant.

If adaptation to dd pattern is associated to a better discrimination ability compared to other patterns, this would have several implications. First, this would indicate that the different effect of dd and Dd patterns is indeed due (at least in part) to non-local adaptation, and not just to a criterion shift, as the effect persists in a perceptual comparison (2AFC) task. Since the local, peripheral density is identical in both dd and Dd, the difference must be related to the central high density, either through non-local physical effects (as seen in the Central session of Experiment 6A) or through the influence of UI-driven changes.

A second implication may be drawn from our results, albeit concerning only the mechanisms of density processing and unrelated to our question about the neural basis of UI. The fact that the direction of the supposed effect is the same in 27a and 27b (i.e. for lower and higher test densities) strongly suggests that, in the conditions of our experiment, density adaptation is unidirectional, i.e. it can only decrease perceived density. Bidirectional adaptation, as observed for channel-based dimensions such as orientation, produces a repulsion in both directions of the perceptual space. Thus, if adaptation to density was bidirectional, we should expect a symmetrical effect, depending on the perceptual distance between adaptor and test: in other words, the effect of dd pattern on the discrimination of lower test densities should be similar to the effect of DD pattern on higher test densities. Rather, it seems that the high adapting density (D, 0.4), either presented on the whole screen (DD) or only in the central area (Dd), affects the perceived density on all test and reference density values (all of them lower than 0.4), in a manner that suggests a perceptual compression of test and reference densities as they are both repelled toward the lower end of the perceptual space. This would account for the decreased discrimination ability after adaptation to a pattern containing high density. On the other hand, the low density (d, 0.2) would not

have any effect whatsoever in case of unidirectional adaptation, because test and reference densities are higher than d. This would explain the greater discrimination, and better performance, in dd blocks compared to any pattern containing high density. Indeed, most studies on density adaptation have reported unidirectional effect (Durgin, 1995; Durgin & Huk, 1997), and those who have found bidirectionality achieved it by implementing specific designs (like sequential 2IFC presentation instead of simultaneous 2AFC task) (Sun et al., 2017).

Thus, our results (even if underpowered and non-significant) suggest that the high-density central area in the pattern suitable for producing UI (Dd pattern) has an effect on density discrimination in the periphery, which is consistent with non-local adaptation, and not merely a criterion shift. However, it is still not clear if this effect is explained by non-local effect of the physical density in the central area or rather due to a local recodification of density information in the peripheral receptive fields, driven by the Uniformity Illusion. This distinction, key to our investigation on the basis of UI, may be addressed by examining time dependency of illusory and physical uniformity in Dd blocks of the Illusion and Control session. Like for Experiment 2A, we can only draw a valid conclusion if we find time dependency in the Control session: in that case, finding it *also* in the Illusion session would mean that UI causes a recodification of peripheral information similar to the physical control, while failing to find it would suggest that the effect seen for the Illusion Dd blocks is due to non-local adaptation by the physical high-density centre. On the other hand, if there was no time-dependency in the Control session, we would not be able to disambiguate between both possibilities.

We therefore computed bivariate correlations between each participant's proportion (%) of time of perceived or physical uniformity (in Dd blocks) and their proportion of correct responses. These analyses were split by session and by test density values lower and higher than density. Time dependency should see results for longer times of uniformity approximate to results observed for DD blocks: i.e., a negative correlation between time of uniformity and proportion of correct reports. Although entirely non-

significant, at first sight there seems to be a negative correlation for both Illusion and Control session; if real, this would indicate an effect of UI on density adaptation. However, on closer inspection, this purported negative correlation appears entirely driven by a single participant who consistently performed at or below chance (<50%). Dismissing this data point, it would seem that response patterns in Dd blocks are time-invariant for the Illusion and Control session, meaning that, as for Experiment 6A, density adaptation is not an adequate paradigm for exploring UI.

CHAPTER 9: DISCUSSION

In this chapter we discuss the results of Experiments 5 and 6 and the possible basis of the Uniformity Illusion (UI). Our findings indicate that perceptual uniformity under UI (on orientation) does not require changes in encoding of peripheral information in the primary visual cortex (V1), while the involvement of other areas and mechanisms is less clear. It is possible that UI represents a heterogeneous collection of phenomena arising at different brain areas for different visual dimensions. The non-local processing of feature information in UI suggests the involvement of texture processing - in line with this, previous studies have not encountered changes in V1 activity in texture filling-in. Furthermore, UI shares some properties (but also large differences) with peripheral crowding, a process that has also been linked to texture processing and does not affect V1-based tilt adaptation. In addition, it is very likely that perceptual inflation (a decisional or metacognitive bias for uniformity) is at least partly responsible for the phenomenology, as suggested by the pattern of UI reports in our experiments. The contents of this chapter have been published in i-Perception journal as part of the article 'The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence from Sensory Adaptation' (Suárez-Pinilla et al., 2018a) .

The Uniformity Illusion (UI) is a striking phenomenon whereby experience across the whole visual field is modified by higher-precision foveal information, so that the peripheral elements of a display appear to uniformly take the properties of the fovea. Intuitively, this illusion seems to bring forth a promising approach to the study of naturalistic visual experience, in particular about surface processing across large visual

angle distances, affected by varying degrees of sensory precision. However, the underlying mechanisms of UI remain unknown.

Using a version of UI with oriented Gabor patches, we found that UI does not produce an orientation adaptation after-effect consistent with the illusory percept. Instead, orientation after-effects only ever followed the (local) physically presented orientation. This suggests that the UI, at least on orientation, arises from higher-level (higher than primary visual cortex) perceptual processes, without recoding of peripheral signals in V1 neurons.

We then employed a similar paradigm applied to a visual property that is processed beyond V1, specifically texture-density. However, physical presentation of different densities (as required for producing UI) drove non-local and time-invariant effects on responses about density. Whether this was due to non-local adaptation to texture-density or to a shift in internal decision criterion, this rendered the experimental paradigm unsuitable for assessing a potential UI-driven effect. We thus designed a modified experiment employing a perceptual comparison (instead of appearance) task, which was more robust to criterion shifts; nevertheless, results were convergent (albeit underpowered), suggesting non-local and time-invariant density adaptation, which prevented disambiguating between adaptation directly driven by the central area and a UI-driven effect due to sensory recodification in peripheral receptive fields. Still, although the absence of such effect couldn't be demonstrated, it is worth noting that every aspect of our results could be explained in light of other demonstrated properties, without need for a hypothetical UI-driven adaptation.

It has been suggested that UI may result from predictive processing operations in the visual hierarchy (Otten et al., 2016). In a hierarchical predictive coding scheme, perception arises from the interaction of bottom-up sensory signals with top-down expectations generated in higher cortical areas (K. Friston, 2005; Rao & Ballard, 1999). Prediction error is determined by the discrepancy between both and propagates

through the sensory hierarchy to update the internal world model. Although the interplay between neural signatures of sensory adaptation and predictive coding is not fully understood (Symonds et al., 2017), evidence indicates that top-down expectations produce activity changes in the visual cortex - also specifically for orientation-selective neurons in V1 (Schummers, Sharma, & Sur, 2005), with adaptation adjusting the relative weight of bottom-up and top-down signals in relation to their precision (Malmierca, Anderson, & Antunes, 2015). Under this framework, UI may be conceptualized as the result of high-precision foveal signals which attain more weight than the periphery in forming perceptual predictions for the presented pattern – possibly in combination with a prior for perceptual uniformity for the entirety of the visual field. After a period of exposure, adaptation renders low-precision peripheral signals still weaker (in a manner resembling Troxler fading (Balas & Sinha, 2007; Martinez-Conde & Macknik, 2017)), until eventually they become unable to overcome the central-based prediction (Otten et al., 2016).

Our results suggest that if UI does result from such predictive operations, the locus of influence of the feedback does not reach primary visual cortex, as illusory uniformity on orientation produced no measurable adaptation effect.

What, then, is the neural basis of UI? UI might be an instance of perceptual filling-in, a phenomenon whereby, a visual attribute like colour, luminance, or texture is perceived in a region of the visual field where it does not physically exist, by virtue of its presence in an adjacent area (Komatsu, 2006). However, unlike typical instances of uniform spread of colour or luminance, in our examples (UI on orientation and texture-density), the distinction between background and grid elements persists and the illusion selectively informs the appearance of the individual Gabors, or the 'clutteredness' of the dots. The process may be similar to texture filling-in or involve texture processing in a broader sense. Notably, several neurophysiological and neuroimaging experiments have reported changes in neural activation in early visual areas that correlate with perceptual filling-in: however, while for colour or luminance this correlate has been seen at V1

(Hsieh & Tse, 2010), for texture filling-in it has only been observed at V2 and above (De Weerd et al., 1995; Komatsu, 2006) - in agreement with our results.

As previously mentioned, some properties of UI greatly echo the mechanism of texture perception, mainly regarding the non-local, statistics-based processing of feature information (S. C. Dakin & Watt, 1997; Julesz, 1981; Thielscher, Kolle, Neumann, Spitzer, & Gron, 2008). It is possible that UI arises when summary statistics in fovea and periphery are similar enough to prevent discrimination into different textures (S. C. Dakin & Watt, 1997). Texture perception appears to depend on areas beyond the primary visual cortex, specifically V2 and V3 (El-Shamayleh & Movshon, 2011; Okazawa, Tajima, & Komatsu, 2017; C. M. Ziemba, Freeman, Movshon, & Simoncelli, 2016).

UI also exhibits similarities with crowding, as a context-dependent alteration of peripheral perception. Like UI, crowding arises for different low and high-level dimensions and at several stages of the visual system, involving V2 and above (Whitney & Levi, 2011) - for instance, tilt adaptation to the veridical orientation is present for crowded, indistinguishable stimuli (He et al., 1996). Crowding has been likened to texture perception (L. Parkes et al., 2001), and it has been reported that texture-density perception depends on crowding-like processes (Anobile et al., 2015). However, as a fundamental difference with crowding, in UI peripheral phenomenology is not a mixture of adjacent stimuli, but the replacement of peripheral appearance by the traits of sometimes distant foveal elements.

Finally, UI may be due to perceptual inflation, whereby apparent detail in the periphery is not sustained on perceptual content, but due to decisional or metacognitive biases (Odegaard et al., 2018). In both of our experiments (orientation and density), during the Control session, where a physically uniform pattern was presented at times, participants were less prone to report UI during presentation of the non-uniform pattern compared to the Illusion session: this suggests a shift in decision criterion for uniformity that determines the experience of UI. Importantly, these processes are not exclusive:

possibly both texture processing and perceptual inflation contribute to UI. Further studies may elucidate the precise contribution of the different perceptual mechanisms that underlie foveal-peripheral integration, as demonstrated by UI, and that are central to naturalistic visual experience. However, our results clearly demonstrate that, at least for orientation, these mechanisms do not alter neural coding at the primary visual cortex.

PART IV: Conclusions

Visual perception relies on contextual information for the processing of sensory signals, as a way to disambiguate signal and noise and to optimize the neural code to the statistics of the environment. A well-known example of this is sensory adaptation, whereby perception is shifted *away* from previous information -or adjacent information, in surround suppression and other extra-classical effects. Beside this negative bias, positive biases *toward* the spatiotemporal context have been recognized for long, and in recent years have been increasingly studied in terms of Bayesian operations, particularly within the predictive coding framework, which regards perception as Bayesian inference. A fundamental trait to gauge the properties of the environment, tune noise management and metacognition, and optimize Bayesian computations is visual variability, which provides a measure of the reliability of the sensory signal (or width of the likelihood distribution) as well as the range of expected stimuli (or width of the prior). In turn, perception of visual variability is also subject to tuning by contextual cues over time and across the visual field, as we have studied in a series of experiments reported in this thesis.

Firstly we studied serial dependence in visual variance, i.e., how judgments about motion variance are influenced by previous variance presentations. We identified two history-dependent biases: a positive, Bayesian-like bias exerted by recent history and of likely decisional origin, and a negative effect compatible with sensory adaptation. The origin of positive serial dependence in perception is a matter of intense debate, with several studies presenting evidence for (Cicchini et al., 2017; Fischer & Whitney, 2014; M. Fornaciai & Park, 2018; John-Saaltink et al., 2016) and against (Alais, Lelung, et al., 2017; Bliss et al., 2017; Fritsche et al., 2017; Suárez-Pinilla et al., 2018b) a perceptual nature. There is, besides, a large body of knowledge about recency biases in working memory that may shed some light on the issue (Akrami et al., 2018; Ashourian & Loewenstein, 2011; Fassihi et al., 2014; Huang & Sekuler, 2014; Lockhead & King, 1983; Lu et al., 1992; Olkkonen et al., 2014; Papadimitriou et al., 2015; Preuschhof et al., 2010; Romo & Salinas, 2003; Visscher et al., 2009). In our experiments, serial dependence in variance was necessarily driven by a past, confident decision about variance and

exhibited superficially similar characteristics to the interaction of successive working memory representations (Huang & Sekuler, 2014; Kalm & Norris, 2017). No positive serial dependence was observed if there had not been a decision about variance (for the same amount of attention during perception, as ensured by post-cued task-switching), or when confidence about such decision had been low. In a continuous flash suppression experiment about serial dependence in orientation, we further confirmed that, for a visual dimension that is mainly processed in the primary visual cortex, serial dependence could not arise just as a result of local changes in V1 neural activity if higher-level processing had been blocked by masking in the past trial. While this last result does not rule out that a change in sensory areas may still happen due to top-down effects, it points to a fundamental difference with another well-known history-dependent sensory bias such as (negative) adaptation to orientation, which persists for unseen or crowded signals. In summary, we propose that perceptual decision-making, and particularly the processing of visual variability, relies on two opposite effects of stimulus history, originating at different levels of processing: a negative sensory effect and a positive decisional bias. Our conclusions agree with those by Fritsche and colleagues about the opposite effects of perception and decision in serial dependence (Fritsche et al., 2017), and with fMRI evidence of different areas involved in perceptual hysteresis (higher-order visual and fronto-parietal areas) and adaptation (early visual areas) (Schwiedrzik et al., 2014). This multi-level combination of contextual effects would help to manage the balance between the needed sensitivity to environmental changes and the maintenance of a stable and smooth experience of the natural world, where changes are relatively discrete and infrequent.

Furthermore, we built a two-layer model based on the premise of the existence of two competing biases occurring at different levels of perceptual decision-making, by utilising standard modelling approaches for attractive, recency-driven biases or regression effects (iterative Bayesian operations) and sensory adaptation effects (population codes). This model was able to replicate our own findings in their approximate timescales, showing a reversal from decisional positive bias with respect to recent

presentations toward a sensory negative effect by the broader context. These two effects may be assimilated to changes in criterion (decisional bias) and signal precision (sensory bias) as defined in models of signal detection theory (Stanislaw & Todorov, 1999).

In a second series of experiments we examined the management of variable information presented across different eccentricities, and how peripheral variability is 'replaced' by a uniform appearance informed by foveal input in the Uniformity Illusion (UI). In a somehow similar manner to our serial dependence results, regarding UI on orientation, we observed a dissociation between low-level negative adaptation and a higher-level bias toward uniformity that did not depend on sensory changes. Specifically, we observed that sensory adaptation (revealed by the V1-based tilt after-effect) and the attractive bias toward contextual (foveal) information that constitutes UI phenomenology were independent effects that may take place simultaneously and did not interact with each other, as the latter depended on higher-level processes. Our findings about UI on density were less conclusive due to the less local character of density adaptation, but, while a sensory effect of UI on density could not be completely ruled out, it proved unnecessary for describing our results. Furthermore, the pattern of UI reports, for both orientation and density, was highly suggestive of a decisional bias as a cause (or large contributor) for UI experience. Thus, participants reported illusory uniformity far less frequently, for an identical non-uniform pattern, when a physically uniform pattern was presented at times, suggesting a conservative criterion shift for UI.

Although the processes underlying perceptual stability over time and across the visual field, as illustrated by serial dependence and UI, are very different in properties, timescales, and probably in their underlying mechanisms, both operate without apparent changes in low-level sensory information, and likely mainly at a decisional level. It is probable that decisional and metacognitive biases have a much greater role than usually recognized in the construction of visual awareness, and particularly in our illusion of living in a fundamentally homogeneous and stable world.

PART V: Additional Research: Modulation of Subjective Time Perception by Perceptual and Physiological Variability

Abstract

The main research topic of this thesis dealt with the modulation of visual variability by the spatiotemporal context, as detailed in Parts I-IV. In addition to this, we also investigated how visual variability (the amount of change in perceptual content over time and across the visual field) determines a high-level property at the core of our conscious experience: namely, the subjective perception of passage of time. A model of time perception based on an artificial perceptual classification network recently demonstrated biases in duration estimates driven by visual content, similar to those of human participants when fed with naturalistic video scenes (Roseboom et al., 2017). This finding supported that variability of perceptual content is the basis of subjective duration. Here we analysed the human participants data employed in the aforementioned study, to contrast the influence of external variability (change in visual content) with variability of internal processes, in terms of determining trial-by-trial differences in duration estimation. Specifically, we aimed to assess the potential effect of autonomic processes (signalled by heart rate and pupil size) and dopaminergic activity (using spontaneous blinking as proxy marker), as both have been deemed central to duration perception according to previous studies. In our data we unveiled an influence of visual content in duration estimates which could not be trivially explained by the correlation of visual scenes with a self-generated signal such as eye movements. However, and contrary to previous claims, neither heart rate, pupil size nor spontaneous blinking showed statistical association with trial-by-trial duration estimates. We propose that, as demonstrated by the model performance in (Roseboom et al., 2017), change in perceptual content underpins human duration perception. In presence of naturalistic visual stimulation (unlike simple laboratory tasks or interoception paradigms), visual variability forms a major part of the multimodal perceptual content, with visceral or neural states showing a

negligible influence in time perception. As mentioned before, Part V is based on the same dataset as (Roseboom et al., 2017), but all the analyses reported here are my own.

CHAPTER 10: INTRODUCTION

The mechanistic basis of subjective duration perception is much debated. Many studies postulate the existence of a neural clock driven by rhythmic or oscillating neural activity, with dopaminergic oscillators featuring prominently in these theories. Other authors propose that the accumulation of visceral states in specific brain areas gives rise to the sense of duration, relating embodied time perception with a pacemaker-accumulator framework. In support of these theories, several studies have sought for an association between accessible body signals reflecting neural or visceral processes and duration estimates, with variable success. Conversely, alternative accounts propose that, rather than internal processes, the basis of duration perception is the change in perceptual content, mainly driven by external stimuli. A recent study has provided support to this theory by modelling human-like biases in duration perception in terms of changes in visual content quantified by an artificial perceptual classification network. We analysed the human data obtained for that study by contrasting the influence of external content and measured body signals in trial-by-trial differences in duration estimates.

Duration perception in the range of seconds to minutes is an essential feature of brain function, critical to the interpretation of environmental events, as well as to learning, strategical planning and decision-making processes (Gallistel & Gibbon, 2000; Meck, 2003). However, the neural mechanisms that give rise to duration perception are still unclear.

Most accounts of how a phenomenological experience of duration arises rely on the assumption that there is a mechanism mapping physical into perceived time, i.e. a neural clock (Matell & Meck, 2004; Treisman, Faulkner, Naish, & Brogan, 1990). Among these, the influential pacemaker-accumulator model proposes that the experience of duration arises from a neural clock wherein a pacemaker generates sequential neural pulses that are stored in an accumulator: according to this, the number of pulses over a certain interval constitutes the brain's estimation of the duration of such interval (Treisman et al., 1990). A variation of this model sees time as a metric arising from the dynamics of multiple neural oscillators with their phasic activity operating on different timescales (Church, 1984; Matell & Meck, 2004; Mauk & Buonomano, 2004).

Evidence presented in favour of the putative role of neural clocks has linked striatal dopamine to duration perception (Melissa J. Allman & Warren H. Meck, 2012; Jennifer T Coull, Cheng, & Meck, 2011; Jennifer T. Coull, Hwang, Leyton, & Dagher, 2012; Matell & Meck, 2004; Meck, 2006). Specifically, increased dopaminergic activity has been related to overestimation of duration, and vice versa, in the milliseconds to seconds range (Jennifer T Coull et al., 2011; Terhune, Sullivan, & Simola, 2016). This has led several researchers to propose a fundamental role for neural oscillators forming part of the ascending nigrostriatal dopamine pathway of the dorsal striatum (Jennifer T Coull et al., 2011; Matell & Meck, 2004; Meck, 2006). Taking advantage of the link between increased striatal dopamine and spontaneous blinking (Karson, 1988), Terhune and colleagues reported transient variations in duration estimation in the sub-second and supra-second range, with human participants systematically biased towards reporting durations as longer immediately after spontaneous blinking than in the absence of a prior blink (Terhune et al., 2016).

By contrast with neural clocks, a basis for human duration perception in physiological, rather than neural processes has long been suggested (Bell & Provins, 1963; Münsterberg, 1899; Osato, Ogawa, & Takaoka, 1995; Pollack, Ochberg, & Meyer, 1965; Schaefer & Gilliland, 1938; Schwarz, Winkler, & Sedlmeier, 2013) and in recent years has

again become popular (A.D. Craig, 2009; Lernia et al., 2018; Meissner & Wittman, 2011; Wittmann, Simmons, Aron, & Paulus, 2010). Such a proposal may be linked with the pacemaker-accumulator model, by making physiological processes the source of the rhythmic pacemaker pulses (Lernia et al., 2018). Along these lines, suggestions focus on the role of rhythmic interoceptive signals - the heartbeat specifically. Although researchers have long attempted to uncover a relationship between cardiac activity and duration perception, most of the attempts have been unsuccessful, or results have been partial or non-straightforward (Bell & Provins, 1963; Osato et al., 1995; Schaefer & Gilliland, 1938; Schwarz et al., 2013). Only few studies have found suggestive changes in cardiac activity during the encoding of time intervals (Meissner & Wittman, 2011; Pollatos, Yeldesbay, Pikovsky, & Rosenblum, 2014), though a major confounding factor to some of these studies is the association of autonomic activity to variations in arousal or attention that may have unspecific effects on time perception (as well as other cognitive) tasks (Schwarz et al., 2013).

Alternative accounts for duration perception suggest that, rather than being (primarily) internally-driven by neural or physiological clocks, the basis for duration perception lies in changes in perceptual content (Herbst, Javadi, Meer, & Busch, 2013; Ryota Kanai, Paffen, Hogendoorn, & Verstraten, 2011; Linares & Gorea, 2015), i.e., in the amount of visual variability across the visual field and throughout time. Under these accounts, measurable stimulus attributes, rather than body signals, would be expected to best correlate with duration estimates. This simple approach has often been dismissed on the basis that changes in internal states, such as arousal or attention, can lead to the exact same perceptual content (stimulus) being reported as different in duration depending on the context in which it is viewed. However, a recent development (Roseboom et al., 2017) of this simple idea suggests that it is not how much the perceptual stimulation changes that is key, but rather how much change occurs in neural activity in perceptual classification networks (in response to changes in perceptual stimulation) what determines duration perception. This distinction allows for the possibility that changes in stimulation can drive changes in neural activity related to

changes in perceptual content, but further, that the interpretation of this neural activity is constrained by the state of the system during estimation, allowing for fluctuations in, for example, attention to time (or other task-dependent features).

To support the idea that changes in activity within perceptual classification networks could underlie human duration estimation, a model of time perception based on an artificial perceptual classification network was recently demonstrated (Roseboom et al., 2017). The model was built to produce duration estimates from the input of videos of natural scenes between 1 and 64 seconds. Model-produced duration estimates were found to be well matched to human-produced estimates made regarding the exact same scenes - including matching the pattern of biases in human reports by content of the scene, with busy scenes filmed while walking around a city judged as longer in duration than those in less busy contexts, such as walking around the countryside, or sitting in an office or café. Accompanying the behavioural reports of apparent duration, this study also recorded where in the scene the human participants were looking (using eye-tracking) and monitored the participants' heartbeat (using blood-volume pulse measurements) throughout the experiment. The results reported in the following chapters come from further analyses conducted on this combination of behavioural reports, eye-movements (including saccades, blinks, and pupil size), and heart rate data.

Unlike most studies on duration perception, which typically use simple stimuli (circles on a screen (Terhune et al., 2016), auditory tones (Meissner & Wittman, 2011), etc), or in some cases specifically interoceptive stimuli (Lernia et al., 2018), our study used complex, naturalistic visual scenes, closer to what constitutes everyday phenomenological experience, of which visual information forms a major part. In combination with the eye-tracking and heartbeat data, this dataset allows an explicit contrast of the potential relative roles of the different proposed contributors to human duration estimation – external (stimulus variability) and internal (potentially striatal dopaminergic activity and fluctuations in heart rate) components. To assess the relative

contributions of these factors, we analysed the participants' data to look for associations between external and internal events, and duration estimates:

1. First, we looked at the relationship between visual content and duration estimation. An association between visual content (scene type: city, countryside, and office or café) and duration estimate has previously been reported for this data (Roseboom et al., 2017); Here, we additionally considered the relationship between visual content and saccadic eye movements, and also, regarding saccades as a content-driven behaviour, between saccades and duration estimates.

2. Second, based on the popular idea and previous results suggesting that autonomic activity is a key contributor to duration estimation (A.D. Craig, 2009; Lernia et al., 2018; Meissner & Wittman, 2011; Wittmann et al., 2010), we examined the relationship of cardiac activity and pupil size with duration estimation.

3. Finally, following the findings of Terhune and colleagues (Terhune et al., 2016) in taking spontaneous blinking as an indicator of fluctuations in striatal dopamine activity, we examined the relationship between pre-trial blinking and duration estimates, particularly in the shorter videos (1 - 2 seconds).

CHAPTER 11: METHODS

In this chapter we present the methodology for our experiment, common to that reported in (Roseboom et al., 2017). Human participants watched several silent videos of naturalistic scenes with different amounts of complexity and 'liveliness.' Their task was to report the estimated duration of each. Eye-tracking and blood-volume pulse recordings were conducted during the experimental session. The available data comprised information on gaze position, pupil size, saccades and blinks, as well as cardiac activity, all of them continuously measured during video presentation, together with behavioural reports (duration estimates) provided at the end of each video.

1. PARTICIPANTS

Participants were recruited from the University of Sussex. All of them were over 18 and reported normal or corrected-to-normal vision. They were awarded with course credits or, alternatively, £5 per hour for their participation. The study was granted ethical approval by the Research Ethics Committee of the University of Sussex.

Fifty-five human participants (40 female, average age 21.4) took part in the experiment. Three were excluded from the analysis because a technical issue prevented successful eye tracking recording. After exclusion of those subjects, the entire dataset totaled 4060 trials.

2. STIMULI

Experimental stimuli were based on videos recorded in the city of Brighton (United Kingdom), the University of Sussex campus, and the surrounding countryside. These videos were recorded with a GoPro Hero 4 at 60 Hz and 1920 x 1080 pixels, and further processed at 30 Hz and 1280 x 720 pixels. The video employed in each experimental trial was extracted from a pseudo-random list of 4290 video fragments, comprising 330 repetitions of each of 13 durations ranging from 1 to 64 seconds (1, 1.5, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64 s). There was no attempt to restrict overlap of frames between different video fragments.

The videos could be classified on three video types in terms of content: 1- outdoors-urban-walking, 2- outdoors-other-walking and 3- indoors-sitting. Video type 1 comprises videos recorded while walking around the city, type 2 includes scenes recorded while walking around the countryside or a leafy campus, and type 3 includes quiet scenes on a café or office. Each video type has a higher (density of) **perceptual content** than the next. As mentioned in the previous paper, we define perceptual content as the amount of visual variability across the visual scene (or within each video frame) and through time (across frames). More complex and livelier videos will have a higher perceptual content (or higher perceptual variability) than static scenes involving inanimate objects. Although this term is used loosely throughout the paper, and not as a hard numeric measure, its validity was confirmed in (Roseboom et al., 2017) by feeding the videos into an image classification algorithm that quantified frame-by-frame rate of change at different hierarchical levels of visual information: from pixel wise to object-based changes.

3. APPARATUS

Experiments were programmed in MATLAB 2012b (MathWorks Inc., Natick, US-MA) using Psychtoolbox 3 and the Eyelink Toolbox, and displayed on a LaCie Electron 22 BLUE II 22" with screen resolution of 1280 x 1024 pixels and refresh rate of 60 Hz. Eye tracking was performed with Eyelink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at a sampling rate of 1000 Hz, using a desktop camera mount. Head position was stabilized at 57 cm from the screen with a chin and forehead rest. Calibration of the eye-tracking system was performed at the beginning of each 20-trial block, using a standard 5-point grid and allowing for a maximal average error of 0.5 degrees of visual angle (dva). We used a velocity threshold of 22 dva/s^2, an acceleration threshold of 4000 dva/s^2 and a motion threshold of 0.15 dva for saccade detection.

4. PROCEDURE

Participants typically completed 80 trials in the 1-hour experimental session, organized in 4 blocks with 20 trials each, though due to time or other constraints some participants completed fewer trials. Immediately after the end of a video, they reported its estimated duration in seconds using a visual analogue scale. The videos assigned to each participant were randomized, and neither their content nor their duration was balanced or kept the same across participants. However, all subjects watched at least one video of each of the 13 video durations -except one who, for logistical reasons, lacked trials with three movie durations.

5. STATISTICAL ANALYSIS

Statistical analysis was conducted on Matlab 2016a (MathWorks Inc., Natick, US-MA), R 3.4.4 (The R Foundation for Statistical Computing, http://www.R-project.org) and JASP 0.8.5.1 (JASP Team 2017).

For Bayesian statistical analyses we employed the default JASP priors: for t-tests, a prior distribution Cauchy(0, $\sqrt{1/2}$); for Pearson correlations, a uniform distribution U(-1,1); for ANCOVAs and repeated-measures ANOVAs, r scale prior width of 0.5 for fixed effects, 1 for random effects and 0.354 for covariates. For Bayesian varying-intercepts, varying-slopes linear mixed-effects models (LMMs, conducted in R with the brms statistical package), we chose a uniform prior distribution over the real numbers for the fixed-effects coefficient and for the standard deviation of the by-participant varying intercepts and slopes, and a LKJ prior with shape parameter $\eta = 2.0$ for the random-effect correlation matrices.

The wording employed for describing the amount of evidence indicated by the Bayes factor corresponds to that suggested by Lee and Wagenmakers (Lee & Wagenmakers, 2013). We consider that evidence in favour of the alternative hypothesis is more than anecdotal when $BF_{10}>3$; conversely, there is more than anecdotal evidence in favour of the null when $BF_{10}<1/3$ (equivalently $BF_{01}>3$).

Further, specific details of each analysis are detailed along with their results in the following chapters.

CHAPTER 12: PERCEPTUAL CONTENT AND DURATION ESTIMATION

In this chapter we analyse the statistical association between video content (as a proxy for perceptual content) and duration estimates, as well as the relationship of both variables with saccadic activity. 'Livelier' video types (i.e. urban compared to rural sceneries, and the latter compared to indoor recordings) yielded longer duration estimates for the same veridical intervals, consistent with perceptual change being the basis of duration perception. There was also a relationship between video type and saccade density (saccades per second), but it was non-monotonical (rural scenes had the lowest saccade density), indicating that the aforementioned correlation between video types and duration estimates was not a trivial effect of duration perception being based directly on tracking one's own saccades. In regard to this, there was not a linear relationship between saccade density and duration estimates; conversely, we did find an association between saccade density and accuracy, probably reflecting saccades as an indication of task engagement.

1. RELATIONSHIP OF SCENE TYPE AND DURATION ESTIMATION

Figure 28a depicts participants' responses per video duration. As expected, veridical duration is directly associated with subjective duration estimates. However, an effect of regression to the mean in responses is suggested by a slope of less than one between veridical and estimated durations: in a varying-intercepts, varying-slopes linear mixed-effects model (random effects grouped by subject) with response as dependent variable

and video duration as predictor, the fixed-effect coefficient for video duration is B=0.679 (95% credible intervals 0.621 - 0.738). In other words, for an increase of 1 second in veridical duration, estimated duration increases only 0.679 seconds.

As mentioned in the Introduction (Part V, Chapter 10), it is known that the amount of change in perceptual content influences duration estimation (Roseboom et al., 2017). Considering this, we hypothesized that videos with greater perceptual content, such as those recorded in urban sceneries, would be perceived as longer than videos of the same duration taking place in the countryside or in a quiet office. Figure 28c presents responses per video duration, split by video type: 1- outdoors-urban-walking, 2- outdoors-country-walking and 3- indoors-sitting. On visual inspection, data seems in line with our hypothesis. The following is an extension on that reported previously (Roseboom et al., 2017).

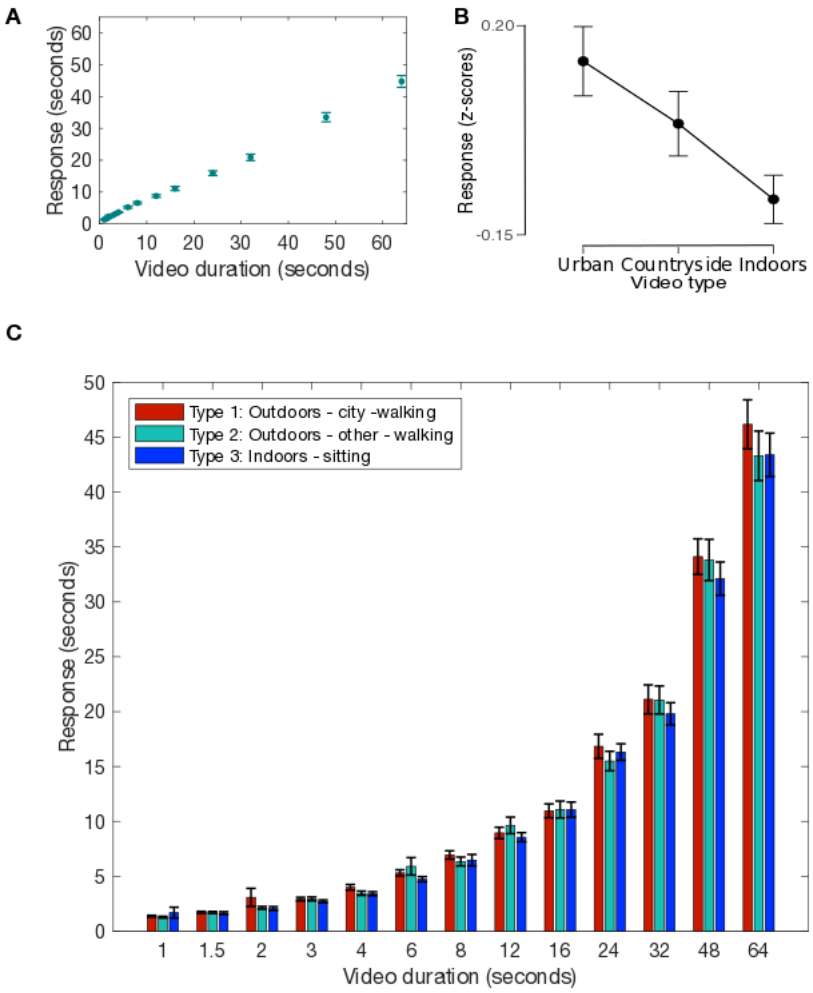


Figure 28. Duration estimation: Behavioural results. 28a. Average response (in seconds) per video duration. The error bars represent between-participant standard error. The slope of less than 1 indicates an effect of regression to the mean: an increase of 1 second in veridical duration produces an increase of less than 1 second in perceived duration. Response dispersion is larger for longer durations, roughly following a power law. **28b.** Normalized response per video type. Type 1: outdoors-city-walking, type 2: outdoors-other-walking, type 3: indoors-sitting. Responses have been normalized according to the distribution of responses provided for each participant per video duration; thus, Z-scores are independent on individual variability and of video duration. Error bars represent 95% credible intervals according to a Bayesian ANCOVA on the effect of video type on normalized responses. The graph shows that subjective duration is positively associated to perceptual content. **28c.** Average response (in seconds) per video duration, divided by video type. The error bars represent between-participant standard error.

To formally test the suggestion that greater perceptual content increases estimated duration in natural scenes, we ran a Bayesian ANCOVA with within-participant normalised response as dependent variable, video type as fixed factor, video duration as covariate and participant's ID as random factor. The best model included video type only (BF_{10}=1.068*10^7). The Bayes Factor for inclusion of video type in the model indicated extreme evidence in its favour ($BF_{inclusion}$=7.136*10^6). Figure 28b depicts the average and 95% credible intervals of the normalized responses per video type: a clear effect is observed, by which video types with greater perceptual content drive longer duration estimates. To ascertain whether this effect happens across all pairwise comparisons between video types, we performed post-hoc analyses and found that evidence for a difference in responses was only anecdotal between types 1 and 2 (both outdoor, walking videos in city and countryside: BF_{10}=1.378), but at least very strong for any of those types compared with type 3 (outdoor versus indoor scenes: BF_{10}=3.621*10^7 for city versus office/café, BF_{10}=34.789 for countryside versus office/café). In summary, our data indicate that videos with greater perceptual content elicit longer duration estimates, supporting the suggestion that subjective perception on the passage of time is (at least partially) driven by perceptual change.

2. RELATIONSHIP OF EYE-MOVEMENTS WITH PERCEPTUAL CONTENT AND DURATION ESTIMATION

2.1. SACCADE DENSITY

The relationship between video type and duration estimates is in agreement with the idea that duration perception is fundamentally related to perceptual change.

The influence of perceptual content on eye-movements is a well-studied area of perceptual science. Both stimulus-driven (especially motion (Itti, 2005; Mital, Smith, Hill, & Henderson, 2011)) and task/context (Henderson, Brockmole, Castelhano, & Mack, 2007; Rothkopf, Ballard, & Hayhoe, 2016; Smith & Mital, 2013) driven factors have been shown to influence eye-movements, with results consistent with minimising prediction error (K. J. Friston, Adams, Perrinet, & Breakspear, 2012; Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Itti & Baldi, 2009; Tatler, Hayhoe, Land, & Ballard, 2011) according to a predictive processing account of perception as Bayesian inference.

Participants in our experiment were not given a specific task in relation to the visual stimuli. It is unclear what influence the more general 'estimate time' task might have on eye-movements. Taking the most naïve position on the possible interaction of eye-movements with our stimuli, city and countryside scenes in our experiment have much more visual motion than the café or office scenes. The city scene might also be expected to have the greatest perceptual novelty as there are many more people and objects continually appearing and disappearing from view as the scene changes. These factors might produce an expected ordering of saccade density (defined as the number of saccades per second) by scene that follows the order for duration estimation reported above: city > countryside > café/office. This pattern of results could imply that an account of subjective time perception based on changes in perceptual content could instead be based on something much more trivial - simply tracking eye-movements - rather than changes in perceptual content itself. Conversely, one might anticipate that for café and office scenes, though the perceptual changes are the most infrequent changes and the scene contains the least amount of motion may contain strong novelty because infrequent changes may be much more salient in their novelty. Below, we investigate the relationship between saccade density and scene type as well as the relationship between saccade density and duration estimation directly.

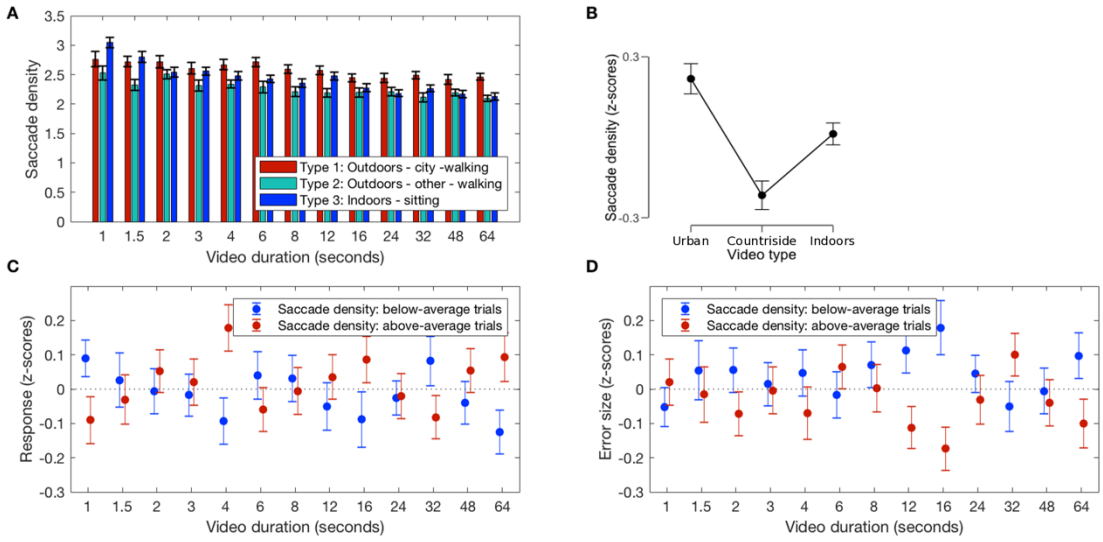


Figure 29. Duration estimation: Saccades and behavioural results. Relationship between video duration, perceptual content, saccades and duration estimates. **29a.** Saccade density by video duration, split by video type. Error bars represent the between-participant standard error. Average saccade density is consistently higher for urban sceneries (for durations over 2 seconds) and decreases with video durations for all video types, but specially for type 3 (static indoor scenes). **29b.** Saccade density (normalized by participant and video duration) by video type: 1: outdoors-urban-walking, 2: outdoors-other-walking, 3: indoors-sitting. Error bars represent 95% credible intervals according to a Bayesian ANCOVA on the effect of video type on saccade density. As expected, saccade density is maximal in urban videos, but its association to perceptual content is not straightforward for the other two video types. **29c-29d.** Relationship between average saccade density and behavioural results. **29c.** Relationship between saccade density and duration estimation. Error bars represent the average response (normalized by participant and video duration) and between-participant standard error, according to video duration (horizontal axis). Each participant's trials corresponding to each video duration are split into two categories, according to whether saccade density was below or above its average for that participant and video duration. Thus, the plot is split according to within-participant (and duration) statistics. For example, regarding 4-second videos, participants on average underestimated video durations in those trials where their saccade density was below their own average for 4-second videos; this underestimation is relative to each participant's own average response for 4-second videos. However, note that each participant watched about 6 videos of each duration: thus, each bar contains information about 3 trials from each of the 53 participants. No clear dependency is seen for most video durations between saccade density and response, making it likely that the mentioned example is a product of multiple comparisons. **29d.** Relationship between saccade density and error size. Error size is calculated as the absolute value of the relative error (error size=|response-duration|/duration), i.e. it represents the amount of deviation from veridical duration, relative to the magnitude of the latter, regardless of the direction of such deviation. Although an overall pattern is not clear, it appears as if participants were more accurate when they performed more saccades in several video durations (particularly clear for 12 and 16-second videos).

*2.*1.1. Saccade density and perceptual content

Figure 29a depicts the average saccade density per video duration and type. Three interesting trends may be observed in relation to the effect of video duration, video type, and their interaction with saccade density. First, consistent with previous results (T. Buswell, 1935; Unema, Pannasch, Joos, & Velichkovsky, 2005), a trend toward lower saccade density for longer video durations can be seen, possibly reflecting a relationship between saccade density and novelty, which would be maximal at video onset and decline in longer videos. Second, for durations over 2 seconds, saccade density is higher in videos walking around city scenes than in the other two, 'less lively' (less variable) types. Third, the aforementioned decrease of average saccade density for longer durations is particularly pronounced for static indoors scenes: this again supports a relationship between saccade density and novelty, since the continuous appearance of new events and objects in the other two video types (recorded while walking around) would ensure a certain amount of novelty throughout the entire duration of the video, unlike in more stationary scenes. In summary, visual inspection of the average saccade density per video duration and type appears to confirm a relationship between saccade density and perceptual content in our data.

To formally test this, we built a Bayesian ANCOVA with saccade density as dependent variable, with video type as fixed factor, video duration as covariate and participant's ID as random factor. A comparison between all possible models was performed: the null model, two models with a single main effect (video duration and video type, respectively), a model with both main effects and the full model with both main effects and their interaction. Participant's ID was part of all competing models, including the null. The most explanatory model for saccade density was the full model with both main effects (video duration and type) and the interaction (duration*type), with a Bayes factor $BF_{10}=1.015*10^{70}$. There was extreme evidence ($BF_{10}>100$) in favour of this model compared to any other competing model, as well as extreme evidence for the inclusion of any of the main effects and the interaction term in the model. This confirmed that

the three aforementioned trends identified by visual inspection of the plots (influence of video duration, video type and interaction) were statistically significant. We performed post-hoc pairwise comparisons between the individual video types: evidence for a difference in saccade density between all possible pairs was also extreme. However, the association was not straightforward as average saccade density was highest for city scenes (type 1), smallest in countryside scenes, and intermediate for office/café scenes -as seen in figure 29b.

To simplify the analysis and isolate the influence of video type (the relevant factor to us at the moment, related to perceptual content of the scenes) we repeated the ANCOVA analysis with saccade density normalized per participant and video duration. The model with the most evidence contained only video type (BF_{10}=8.149*10^{22}), indicating extreme evidence for an effect of video type on saccade density. We performed post-hoc pairwise comparisons between the individual video types: evidence for a difference in saccade density between all possible pairs was also extreme (BF_{10}>100). However, the relationship between saccade density and duration estimation was not straightforward, as average saccade density was highest for city scenes, (Figure 2b).

These results rule out trivial interpretations of the association of change in perceptual content (indicated by video type) and duration estimation, as reported above and supported by modelling work in (Roseboom et al., 2017): duration estimation based on tracking eye-movements alone would not produce the observed pattern of biases in duration estimation as the saccade density is non-monotonically related to duration estimation (compare Figures 28b and 29b).

2.1.2. Saccade density and duration estimates

Having established a relationship between video type and duration estimates (Figure 28b), and now also between video type and saccade density (Figure 29b), we next

enquired whether there would be a direct association between saccade density and duration estimates. Eye-movements have a potentially complicated relationship with time perception. Temporal distortions on short time scales have been reported around the time of saccades in several contexts (D. Burr & Morrone, 2006; Yarrow, Haggard, Heal, Brown, & Rothwell, 2001), although these can sometimes be reduced to changes in apparent content/intensity (Terao, Watanabe, Yagi, & Nishida, 2008). As in eye-movements, both visual motion (Au, Ono, & Watanabe, 2012; brown, 1995; R. Kanai, Paffen, Hogendoorn, & Verstraten, 2006) and novelty have been demonstrated to have a strong influence on duration estimation; the latter on shorter timescales through regularity/oddity effects (Chang, Schwartzman, VanRullen, Kanai, & Seth, 2017; Di Luca & Rhodes, 2016; Eagleman & Pariyadath, 2009; Horr & Di Luca, 2015; Tse, Intriligator, Rivest, & Cavanagh, 2004) and on longer timescales through stimulus/sequence complexity and contextual change (Block, 1982; Block & Reed, 1978; Ornstein, 1969; Poynter, 1983; Poynter & Homa, 1983; Zakay, Tsal, Moses, & Shahar, 1994). Here we examine the association between density of saccadic eye-movements and duration estimation across our measured intervals of 1-64 seconds.

Figure 29c presents participants' average normalized response per video duration, split into trials with below and above-average saccade density, compared with other trials from the same participant and with the same duration. Overall, there is no clear pattern revealing a trend for under or overestimation of duration as a function of saccade density, not for all video durations or for a narrower range of intervals.

To formally test this, we ran a Bayesian ANCOVA with normalized response as dependent variable, normalized saccade density and video duration as covariates, and participant's ID as random factor. Normalization was performed for participant and video duration. The best model included only normalized saccade density, but, with a BF_{10}=1.091, it was barely more likely than the null model (the reference model containing only ID: BF_{10}=1.000). Thus, there was virtually no evidence in favour of (or against) an effect of saccade density on duration estimates, when considering the entire

sample – a result consistent with the non-monotonic relationship between saccade density and video type reported above.

We further assessed whether an association between saccade density and responses could be present only for certain video durations and split the dataset according to the latter variable. We ran a Bayesian ANCOVA for each of the 13 resulting datasets, with normalized response as dependent variable, ID as random factor and saccade density as covariate. Only for the longest duration (64 seconds) did we find evidence for an effect of saccade density on response (BF_{10}=23.827) -a positive correlation-, although the significance of this isolated finding is unclear.

In summary, our data suggests a relationship between changes in perceptual content and duration estimates, as well as (albeit less straightforward) between perceptual content and saccade density. However, there is not a clear global effect of saccade density on duration estimates.

2.1.3. Saccade density and accuracy

We then enquired whether saccade density might not be associated with the magnitude of duration estimates, but to their accuracy. In other words, it might drive no systematic bias toward under or overestimation of time intervals, but affect the response precision, or the amount of deviation from veridical magnitude regardless of direction. Saccade density might correlate with a greater attention and engagement with the presented video and consequently with a greater accuracy (or smaller error size).

For quantifying error size we calculated the absolute value of the relative error with respect to veridical magnitude: error size=|response-duration|/duration. Figure 29d presents normalized error size for each video duration, split into trials with below or above-average saccade density. Although an overall effect is not clear on visual

inspection, for several video durations there appears to be a trend toward lower error sizes (greater accuracy) in trials where saccade density was above average.

We formally tested the effect of saccade density on accuracy by running a Bayesian ANCOVA on normalized error size, with normalized saccade density and video duration as covariates and participant's ID as random factor. The best of all possible models according to this analysis contained saccade density as predictor of error size (BF_{10}=49.784); overall, evidence for inclusion of saccade density in the model was very strong ($BF_{inclusion}$=34.247). To ascertain the direction of the effect, we ran a varying-intercepts, varying slopes Bayesian linear mixed-effects model on the effect of normalized saccade density, duration and their interaction on normalized error sizes, with random effects grouped by participant. The fixed-effects coefficient for saccade density was B=-0.040 (95% credible intervals: -0.085 - 0.005); its negative sign indicates an inverse association with error size, as would be expected if greater saccade density was a sign of a greater engagement during those trials. According to our proposed explanation, such effect on accuracy does not necessarily imply a link between saccades and the mechanistic basis of time perception, since it could be an unspecific effect on task performance.

CHAPTER 13: AUTONOMIC SIGNALS AND DURATION ESTIMATION

A long-standing claim about duration perception postulates that it is based on the accumulation of rhythmic physiological signals such as heartbeats; however, evidence for a statistical association has been elusive. A recent study (Meissner & Wittman, 2011) reported slowing-down of heartbeats during encoding of time intervals, suggesting, according to their authors, an accumulating pattern of parasympathetic activity that may act as a clock-type mechanism. We therefore sought for statistical associations between trial-by-trial variability in body signals revealing autonomic tone (specifically heart rate and pupil size) and duration estimates in our data. Neither average heart rate nor pupil size were associated to duration estimation. We did find a progressive slowing-down of heartbeats during the first 5-10 seconds of video presentation, which correlated with the intensity of rapid pupil contraction that was observed at video onset (1-2 seconds), but none of them were in turn associated to duration estimation. We propose that, in our data, these phenomena reflect parasympathetic activation in relation with novel content (Bradley, 2009), but are not directly related to the mechanism of time tracking.

As mentioned in the Introduction to this study (Part V Chapter 10), it has long been suggested that accumulation of repetitive physiological signals may form the basis of human time perception. Craig (A.D. Craig, 2009) postulated that the sequential accumulation of interoceptive afferent information about visceral states in the anterior insula gives rise to the subjective experience of time. The insular cortex has been related to reception and integration of autonomic signals and to interoceptive awareness (A. D.

Craig, 2009; Nguyen, Breakspear, Hu, & Guo, 2016). Following this lead, in a fMRI study, Wittman and colleagues (Wittmann et al., 2010) identified a pattern of accumulating neural activity in the bilateral *posterior* insula during the encoding of time intervals in the range of seconds. It was suggested that this accumulation corresponded to a clock-type pattern, recording sequential physiological states during the attended interval (Wittmann et al., 2010). This proposal connected the pacemaker-accumulator framework to an embodied model of time perception (Lernia et al., 2018), giving support to the intuition that our brain tracks the passage of time on the basis of our own body's signals. However, such intuition greatly predates this theoretical framework.

For many decades, researchers have searched for a link between internal body signals, especially heartbeats, and duration perception. This has been investigated mainly by two approaches: by examining the correlation between heart rate and duration estimates or by inducing experimental manipulations on cardiac activity and assessing duration perception under such conditions. Most studies following the first approach have encountered no correlation between average heart rate and duration estimation (Bell & Provins, 1963; Schaefer & Gilliland, 1938), or obtained weak and partial results (Lediett & Tong, 1972; Osato et al., 1995). As for the second approach, some studies reported distortions in duration perception after inducing the desired physiological conditions by apnoea (Thibaud Jamin et al., 2004), physical activity (Lambourne, 2011; Vercruyssen, Hancock, & Mihaly, 1989) or drugs (G.R. Hawkes, Joy, & Evans, 1962; Jared R. Tinklenberg, Walton T. Roth, & Bert S. Kopell, 1976). A major obstacle to interpretation of results in these studies is that average heart rate also correlates with other factors that could influence task performance in a dimension and task non-specific manner, such as attention or arousal. Studies independently manipulating heart rate and arousal have reported no specific effect of the former on duration estimates (Dormal, Heeren, Pesenti, & Maurage, 2017; Schwarz et al., 2013).

Recently, several authors have proposed that the relationship between cardiac activity and duration perception may be less linear than initially thought. An investigation on

heartbeat synchronization at the onset and ending point of presented time intervals reported marginally significant synchronizations for durations within the range of seconds (3-25 s), but found no association with duration estimates or accuracy (Pollatos et al., 2014). Another study (Meissner & Wittman, 2011) specifically examined the evolution of cardiac activity throughout the encoding of time intervals and found a progressive increase in cardiac periods up to the end of the interval. Furthermore, individuals' duration reproduction accuracy correlated positively both with the slope of cardiac slowing down during the encoding of time intervals and with interoceptive accuracy measured in terms of heart perception scores. The authors reasoned that such slowing down of heart rate would correspond to an accumulation of parasympathetic activity in brain areas responsible for the reception of autonomic input (particularly the posterior insula) that would act as a clock-type mechanism for encoding durations within the range of seconds.

In light of these claims, we looked for similar associations between cardiac activity and duration estimation or accuracy in our own study. We assessed the relationship with both average heart rate and progression of heart rate during the presentation of the videos. In addition, we also examined pupil size as another marker of autonomic activity. Information on interoceptive accuracy or awareness was not available in our experiment.

1. RELATIONSHIP BETWEEN CARDIAC ACTIVITY AND DURATION ESTIMATION

In our study, we obtained blood-volume pulse measurements from participants during video presentation. Heart rate was estimated by using a peak-detection function on blood venous pressure data.

1.1. MEAN HEART RATE

We calculated the mean heart rate per trial (during video presentation) as the ratio of the number of peaks and the video duration, converted to beats per minute.

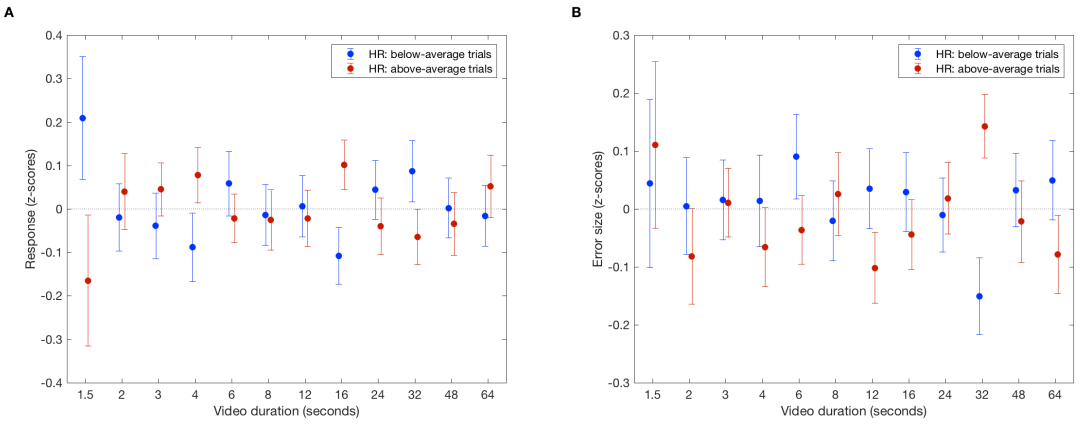


Figure 30. Duration estimation: Mean heart rate and behavioural results. Relationship between mean heart rate (computed trial-wise) and duration estimates. **30a - 30b**. Normalized response (**30a**) or error size (**30b**) by video duration, split into trials with below or above-average heart rate compared to other trials of the same participant and duration.

1.1.1. Mean heart rate and duration estimates

Figure 30a presents the average normalized response by video duration, split into trials with below or above-average heart rate compared to other trials of the same participant and duration. No clear pattern is evident.

The trial-wise relationship between mean heart rate and response was formally assessed by a Bayesian ANCOVA with normalized response as dependent variable, normalized mean heart rate and video duration as covariates, and participant's ID as random factor. As previously, normalization was performed by participant and video duration. According to this analysis, the best model for prediction of (normalized) response was the null model (containing only participant's ID); the Bayes factor for inclusion of mean heart rate indicated very strong evidence against it ($BF_{inclusion}$=0.029), and extreme evidence against the interaction video duration*mean heart rate ($BF_{inclusion}$=$8.136*10^{-4}$).

1.1.2. Mean heart rate and accuracy

Finally, we enquired whether heart rate may produce not a bias in duration estimates, but an effect on its accuracy. We defined error size as the absolute value of the relative error (|response - duration|/duration) and normalized it per participant and video duration: these Z-scores therefore indicate how inaccurate (regardless of the direction of the error) was a participant in a certain trial, compared with their performance in other trials of the same duration.

Figure 30b presents the normalized error size by video duration, split by trials with below and above-average heart rate, respectively. We ran a Bayesian ANCOVA with normalized error size as dependent variable, normalized heart rate and video duration as covariates and participant's ID as random factor. The null model (containing ID) had

the most evidence of all possible models, and the Bayes factor for inclusion of mean heart rate or its interaction with video duration indicated strong (BF_{10}=0.035) and extreme (BF_{10}=0.002) evidence against them, respectively.

In summary, in our data neither duration estimates nor its accuracy were associated with trial-by-trial mean heart rate.

1.2. HEART RATE PROGRESSION

We calculated cardiac period progression through video presentation as described by Meissner and Wittman in (Meissner & Wittman, 2011): cardiac (inter-peak) periods were resampled at 5 Hz using cubic interpolation, averaged on a second-by-second basis and normalized per participant. The slope (linear regression coefficient) of the time series of cardiac period progression indicated the progression of heart rate throughout each trial's video presentation: a positive slope implied that heart rate progressively slowed down (inter-peak periods increased with seconds since video onset) and vice versa.

1.2.1. Heart rate progression throughout video presentation

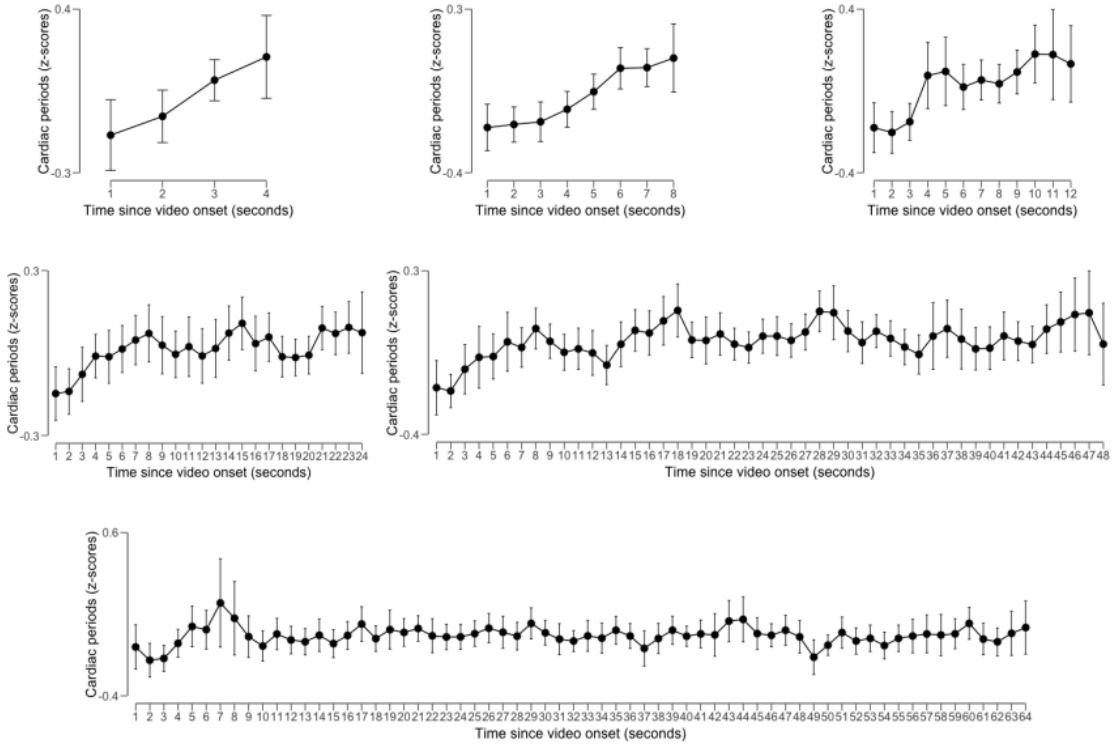


Figure 31. Duration estimation: Cardiac period progression. Average second-by-second cardiac periods for six video durations: 4, 8, 12, 24, 48 and 64 seconds. The error bars represent the 95% credible intervals according to a Bayesian RM ANOVA on the effect of time since onset in cardiac periods. An ascending slope can be observed during the first 5-10 seconds in all cases, indicative of a slowing down of heart rate at the beginning of video presentation, although for long durations a plateau is reached later in the trial.

After obtaining normalized second-by-second cardiac periods for each trial, as described above, we split the dataset into 11 video durations - only considering durations over 2 seconds since second-by-second progression was under examination (see Figure 31). We tested whether there was any effect of time since video onset on cardiac periods by two methods: linear regression and Bayesian repeated-measures ANOVA.

First, we performed a linear regression on the time series of normalized cardiac periods, averaging the time series of each participant and video duration (≥ 2 s). On average, participants' regression slopes (B coefficients) for each video duration were positive in all cases (except for duration of 3 seconds), indicating a slowing-down of heart rate throughout video presentation. To test whether these slopes were truly different than zero, we performed a Bayesian one-sample T-test on participants' mean slopes for each video duration. There was at least moderate evidence in favour of the alternative hypothesis (i.e. $BF_{10} > 3$ for non-zero slopes) for 8 s (BF_{10}=237.363), 12 s (BF_{10}=5.412), 24 s (BF_{10}=8.483), and close to moderate evidence for 4 s (BF_{10}=2.788). Conversely, there was at least moderate evidence in favour of the null ($BF_{10}<1/3$) for 2 s, 3 s, 6 s and 64 s. However, regarding long durations, evidence for an overall slope may be reduced due to the slowing down of heartbeats reaching a plateau after the first 5-10 seconds (see lower panels in Figure 31).

This suggestion of a slowing-down of heart rate throughout the reproduction of the video is consistent with the results reported by Meissner and Wittman during the encoding phase of time intervals of 8, 14 and 20 seconds (Meissner & Wittman, 2011). Considering similar durations in our own data, on average, cardiac periods increased in 0.05, 0.03, 0.01 and 0.01 z-scores per second for 8, 12, 16 and 24 seconds, respectively.

We additionally approached the evolution of cardiac periods throughout time by computing a Bayesian repeated-measures ANOVA on the effect of time since video onset on second-by-second cardiac periods. We ran a separate analysis for each video duration. The dependent variable was the average normalized cardiac period measured at each second since video onset. The only within-subject factor was time (in seconds) since video onset. Thus, the ANOVA had as many levels as seconds of video duration. Figure 4 presents the average cardiac periods and 95% credible intervals at each time point (seconds since onset) for six different video durations. Similar to the previous analysis, evidence for an effect of time since video onset on second-by-second cardiac

periods was at least moderate (BF_{10}>3) for 4 s, 8 s, 12 s and 48 s, with BF_{10}=23.204, 95589.424, 229.741 and 10.545, respectively. On the other hand, there was at least moderate evidence for the null (no different cardiac periods at different time points, with BF_{10}<1/3) for 2 s, 3 s, 6 s, 16 s, 32 s and 64 s. For the intervals similar to those considered by Meissner and Wittman (8, 12, 16 and 24 s), the F values for repeated measures ANOVA were, respectively and corrected by Greenhouse-Geisser method: for 8 s, F= 6.843 (df=3.299, p<0.001, η^2_p=0.118); for 12 s, F=3.715 (df=3.174, p=0.011, η^2_p=0.068); for 16 s, F=1.529 (df = 5.724, p=0.172, η^2_p=0.029); for 24 s, F=1.994 (df=10.49, p=0.029, η^2_p=0.038).

As stated above, for long video durations, the slowing down of cardiac periods reaches a plateau after the first 5-10 seconds, possibly related to a greater attention and arousal at the start of the video, which stabilizes afterward. This stabilization might be the reason why weak or no evidence in favour of the alternative hypothesis is found for the longest durations.

1.2.2. Heart rate progression and duration estimates

In the previous section we reported a reduction in heart rate throughout video presentation. This finding is in agreement with the result reported by Meissner and Wittman (Meissner & Wittman, 2011). These authors also found a positive correlation between heart rate reduction and accuracy in reproduction of the presented interval. They hypothesized that the observed reduction in heart rate up to the end of the interval was consistent with a clock-type accumulating mechanism related to an increase in parasympathetic activity that would track duration perception.

While there were differences between studies concerning the stimulus (an auditory tone versus a naturalistic video) and the task (reproduction versus magnitude estimation), we sought to ascertain whether the observation of a slowing heart rate

during video presentation in our experiment was also associated with duration estimation.

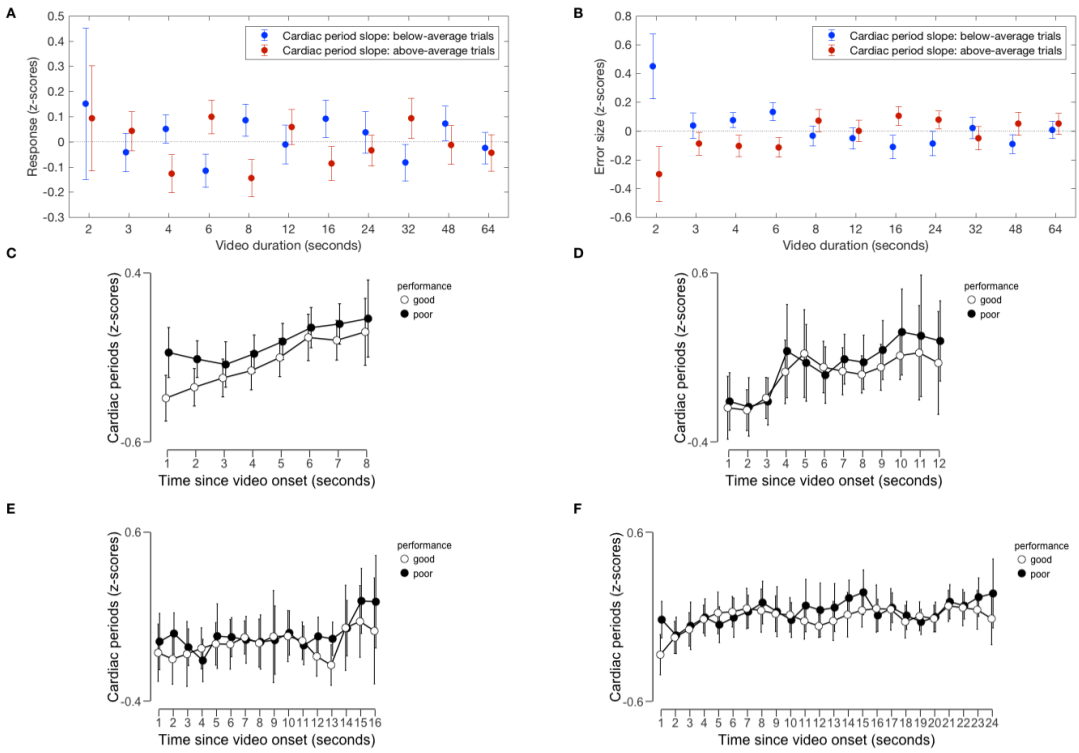


Figure 32. Duration estimation: Cardiac period progression and behavioural results. Figures **32a** and **32b** assess the relationship on a trial-by-trial basis, whereas **32c-32f** compare between participants. **32a-32b**: Normalized response (**32a**) or error size (**32b**) by video duration, split into trials with below or above-average cardiac period slope compared to other trials of the same participant and duration. An above-average slope indicates a more pronounced slowing-down of heart rate throughout video presentation, and vice versa. No clear pattern is seen regarding responses; as for error sizes, on visual inspection there might be an indication of greater accuracy in relation to more pronounced slowing down of heart rate, only for short durations (up to 6 seconds): in other words, for those durations during which the slowing down takes place before reaching a plateau. This might be an indication of greater engagement with the video in the trials where the slowing down is more pronounced, although the supposed association is not clear enough and requires further analysis. **32c-32f**: Second-by-second cardiac periods and 95% credible intervals as a function of time since video onset, for durations of 8, 12, 16 and 24 seconds, split by good and poor performers. Performance classification was made by calculating the average error size (regardless of direction) of each participant and dividing participants according to whether their average error size was above or below the sample median. Thus, in this case the division is made across participants instead of across trials of the same participant. No clear difference in cardiac period progression is seen for good and poor performers (and indeed the results of the RM ANOVA rule out such effect), although, on visual inspection, there might be a hint of a more pronounced ascending slope in good

performers, only during the first 3-5 seconds of the video, for 8, 16 and 24-second videos. If this is the case, it might be indicating a greater engagement at video onset which is related to a better performance, but the fact that the difference (if any) only affects the very first seconds and does not extend up to the end of the video seems in opposition to any purported accumulating clock-type pattern, pointing toward unspecific attention effects.

Figure 32a presents the average normalized response for each video duration, split between trials with below or above-average cardiac period slope compared to other trials by the same participant and duration. As with mean heart rate, no clear pattern is observed that might suggest an effect of heart rate progression on duration estimates.

We ran a Bayesian ANCOVA with normalized response as dependent variable, normalized cardiac period slope and video duration as covariates and participant's ID as random factor. According to this analysis, the model with the most evidence was the null model. The Bayes factor for inclusion of cardiac period slope as predictor of response indicated moderate evidence against it ($BF_{inclusion}$=0.113), and extreme evidence against inclusion of the interaction term cardiac period slope*video duration ($BF_{inclusion}$=0.003).

In summary, heart rate progression throughout video presentation was not related to duration estimation in our study.

Heart rate progression and accuracy

Subsequently we questioned whether, as reported in Meissner's study (Meissner & Wittman, 2011), there was a relationship between heart rate progression and accuracy. We performed analogous analyses as for duration estimation reported above, but employed the error size as dependent variable, defined as in previous sections. The hypothesis in this case was for an inverse association between cardiac period slope and error size.

We approached the relationship between heart rate progression and accuracy in two ways: First, by analysing trial-wise associations within each participant, and second, by examining a possible difference in participant's average cardiac period slope between good and poor duration estimators.

Trial-by-trial association between cardiac period slope and accuracy

Figure 32b shows the normalized error size by video duration, split into trials with below and above-average cardiac period slope. At glance, there appears to be a trend to lower error sizes (greater accuracy) in those trials with above-average cardiac period slope (i.e. steeper slowing down of heart rate), only for short durations (2-6 seconds). Intriguingly, intervals from 2-6 seconds correspond almost entirely with the ascending slope in cardiac period progression, which appears to flatten after 5-10 seconds (see Figure 31). Possibly, the observed heart rate reduction during the first seconds of video presentation might be a sign of arousal and engagement with a new video/task - a phenomenon described before in relation to novel, arousing visual stimuli (Bradley, 2009).

To formally test the existence of any effect of cardiac period progression on trial-by-trial accuracy we ran a Bayesian ANCOVA with normalized error size as dependent variable, normalized cardiac period slope and video duration as covariates and participant's ID as random factor. According to this analysis, the null model outperformed all others, and the Bayes factor for inclusion of cardiac period slope or the interaction term (slope*duration) showed very strong ($BF_{inclusion}$=0.030) or extreme ($BF_{inclusion}$=0.005) evidence against them, respectively. Thus, these results ruled out any effect of cardiac period slope on accuracy on the overall dataset.

We additionally split the dataset by video duration (11 levels) in order to ascertain whether such an effect on performance was only present for certain durations.

However, again we found no support for the inclusion of cardiac period slope as a predictor of performance, even for the short 2-6 second intervals: e.g. $BF_{Inclusion}$ = 0.336, 0.436, 0.347, 0.180 for 2-6 seconds, respectively.

Comparison of average cardiac period slopes in good versus poor performers

Although we failed to find an association between heart rate progression and accuracy in duration estimation on a trial-by-trial basis, we enquired whether this association could be present at the participant, rather than the trial level. In other words, whether there could be individual differences in terms of heart rate progression throughout the encoding of a time interval, which would help to predict a participant's accuracy in time estimation. Specifically, the hypothesis was that good performers (individuals who were more accurate in estimating duration) would present, on average, a more positive slope in cardiac periods (i.e. a more pronounced slowing down of heart rate).

Thus, we followed Meissner (Meissner & Wittman, 2011) and classified participants into good or poor performers, depending on whether their average error size was below or above the sample median. We performed this classification over all video durations pooled, as well as to each video duration separately (according to the latter, one participant could be above the sample's median performance for estimating intervals of 2 seconds, and below the median for 64 seconds).

In a similar manner as when examining the *existence* of a progression in cardiac periods throughout video presentation, here we ascertained the *difference* in such progression between good and poor performers, by two methods: analysing cardiac period slopes and by a Bayesian repeated-measures ANOVA.

First, we performed a Bayesian independent samples t-test on the individual average cardiac period slopes per video duration, with performance as grouping variable. Thus,

we performed two independent samples t-tests per video duration, using either the overall or the duration-specific binary performance classification as grouping variable.

In none of the analyses did we find evidence for the alternative hypothesis (difference in slopes between good and poor performers), and in most cases t-tests supported the null with at least moderate evidence ($BF_{10}<1/3$ or equivalently $BF_{01}>3$). Specifically, considering similar durations as those employed in (Meissner & Wittman, 2011), i.e. for video durations of 8, 12, 16 and 24 seconds, the evidence in favour of the null ($BF_{01}=1/BF_{10}$) was as follows: For 8 seconds, $BF_{01}= 2.575$ for classification based on overall performance, $BF_{01}= 2.178$ based on duration-specific performance. For 12 seconds, $BF_{01}=3.246$ with overall classification, $BF_{01}=2.501$ with duration-specific classification. For 16 seconds, $BF_{01}= 3.324$ with overall classification, $BF_{01}=3.563$ with duration-based classification. For 24 seconds, $BF_{01}= 3.490$ with overall classification, $BF_{01}=3.555$ with duration-based classification.

Subsequently we analysed the effect of time since video onset on second-by-second cardiac period progression in good and poor performers, by considering each participant's time series of average second-by-second cardiac periods for video durations 8, 12, 16 and 24 seconds. We performed two Bayesian repeated-measures ANOVA for each duration, using either overall or duration-specific performance classification as between-subjects factor. The dependent variable was the average cardiac period at each time point (seconds since video onset). We employed time since onset as the only within-subject factor (with as many levels as seconds of video duration) and the binary classification on performance as between-subject factor. If there was any difference in cardiac period progression between good and poor performers, we should find evidence for the inclusion of the interaction term (time since onset*performance) in the model.

In all of the analyses, the best model was either the one containing only time from onset or the null model. This indicated that in some cases (specifically for 8 and 12 seconds)

there was an effect of time since video onset on second-by-second cardiac periods, while in other cases (16 and 24 seconds) there was no evidence for a time-related difference. In any case, neither the cardiac periods at each time point (main effect) nor their progression throughout the trial (interaction with time) were different between good and poor performers. The Bayes factor for inclusion of the interaction term (the one relevant to our question) was always at least moderately against its inclusion ($BF_{inclusion}<1/3$), and extremely against it ($BF_{inclusion}<1/100$) in all cases except for 8 seconds.

Figures 32c-32f present the progression of cardiac periods throughout video presentation for of 8, 12, 16 and 24 second-videos, split by overall performance. The similar progressions observed in good and poor performers confirm that heart rate progression and accuracy in duration perception are not associated in our study. On visual inspection, there might be a more pronounced ascending slope in good performers, only during the first 3-5 seconds of the video, for 8, 16 and 24-second videos. This might indicate a greater engagement at the beginning of the video in good performers, but the fact that the difference (if any) only affects the very first seconds, and does not extend up to the end of the video, opposes the proposal of an accumulating clock-type role involving cardiac period. Rather, these results suggest an unspecific attentional effect on performance, without any specific relation with time estimation.

2. RELATIONSHIP BETWEEN PUPIL SIZE AND DURATION ESTIMATION

In the previous section we reported a progressive slowing down of heart rate that takes place during the first 5-10 seconds of video presentation. We surmise that this is an indication of changes in autonomic activity, probably in relation with an increase in arousal due to engagement with a new video and task (Bradley, 2009), with no evidence of any specific relation to duration estimation. In relation to this, we decided to examine another physiological measurement in our data which is also related to autonomic activity as well as to arousal and visual attention, namely pupil size. Similarly to heart rate, we examined both the mean pupil size during a given trial as well as pupil size progression throughout video presentations.

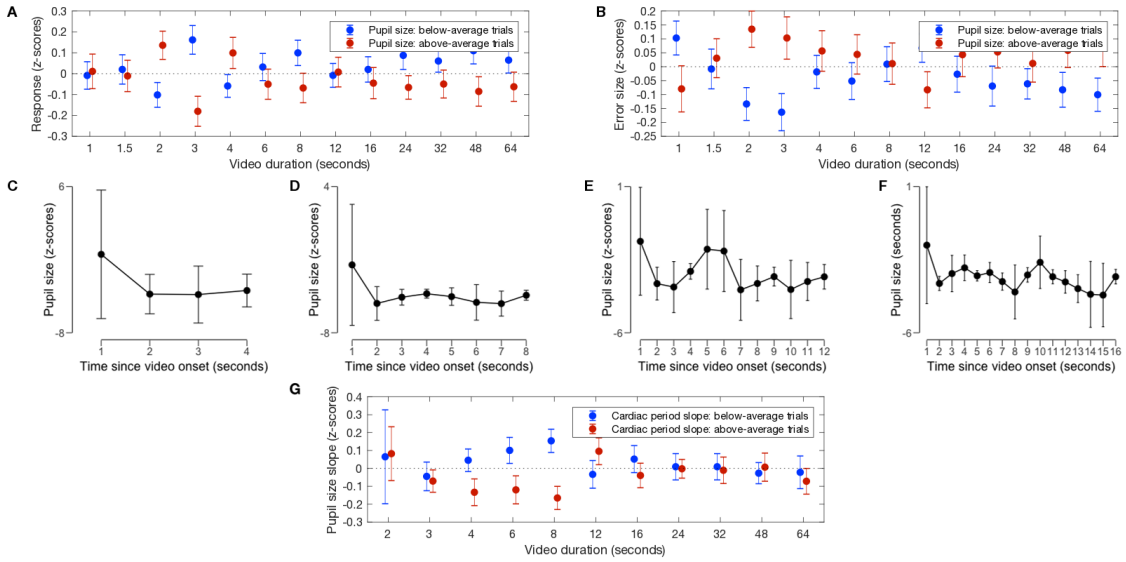


Figure 33. Duration estimation: Pupil size and behavioural results. 33a-33b. Normalized responses (**33a**) and error sizes (**33b**) by video duration, split into trials with below and above-average pupil size compared to other trials by the same participant and duration. At glance, larger pupil size appears to be associated to larger error sizes for several video durations, and possibly also to a relative underestimation of long durations -which might be an indication of regression to the mean effects in trials with lower accuracy. The significance of this supposed relationship with average pupil size is not clear. At any rate, analyses are not statistically significant. **33c-33f.** Normalized second-by-second pupil size as a function of time since video onset, for 4,8,12 and 16-second videos. There appears to be a great variation between participants concerning the average pupil size at video onset, with respect to each participant's average pupil size at all times. This is shown by the large error bars for normalized pupil size at 1 second since video onset, and is probably indicating a large difference in terms of initial pupil response between participants. Despite the large variability, on average there is a trend for a rapid pupil contraction between 1 and 2 seconds since onset, which stabilizes later on. **33g.** Within-participant trial-wise association between cardiac period slope and pupil size slope. The graph presents the normalized pupil size slope by video duration, split into trials with below and above-average cardiac period slope compared to other trials by the same participant and duration. For short video durations (those in which the initial response makes a meaningful contribution to the overall slope), there appears to be an association between larger cardiac period slopes (i.e. steeper slowing down of heart rate) and more negative pupil size slopes (i.e. more pronounced pupil contraction). These two processes probably reflect a parasympathetic activation at video onset, in relation to arousal. Visual inspection of the data (on pupil and cardiac activity) appears to suggest that this initial response is associated with better performance, although statistical analyses do not confirm such trend.

2.1. MEAN PUPIL SIZE

Pupil size was obtained from eye-tracking data. Pupil size of all recorded samples during every video presentation were averaged to generate trial-wise mean pupil size.

2.1.1. Mean pupil size and duration estimates

Figures 33a and 33b present normalized responses (33a) and error sizes (33b) by video duration, split into trials with below and above-average pupil size compared to other trials by the same participant and duration. On visual inspection, there appears to be a positive association between pupil size and error size, as well as a trend toward relative underestimation of long durations in relation to larger pupil size.

To assess whether there was indeed any relationship between pupil size and duration estimates, we ran a Bayesian ANCOVA with normalized response as dependent variable, normalized pupil size and video duration as covariates, and participant's ID as random factor. According to the analysis, the null model was the best predictor of response, and there was strong evidence against the inclusion of pupil size ($BF_{inclusion}$=0.067) and extreme evidence against the interaction video duration*pupil size ($BF_{inclusion}$=0.009).

2.1.2. Mean pupil size and accuracy

Subsequently we tested whether average pupil size may be related to accuracy by running a Bayesian ANCOVA with normalized error size as dependent variable, normalized average pupil size and video duration as covariates and participant's ID as

random factor. The best model was the null (including ID as random factor), and the Bayes factor for inclusion of the average pupil size or its interaction with video duration provided very strong ($BF_{inclusion}$=0.029) or extreme ($BF_{inclusion}$=0.005) evidence against it, respectively.

2.2. PUPIL SIZE PROGRESSION THROUGHOUT VIDEO PRESENTATION

Pupil size progression throughout the trial was assessed by two methods, analogous to those employed with heart rate progression: linear regression slope of the time series of second-by-second pupil size for each participant and video duration, and Bayesian RM ANOVA on the effect of time since video onset on second-by-second pupil size. Our hypothesis was that the participant's pupil would be dilated at video onset and slowly contract afterward. We also hypothesized that this contraction would be correlated with slowing down of heart rate, both of them representing a parasympathetic response during the first seconds of video presentation: thus, there should be an inverse correlation between the linear regression slope for second-by-second cardiac periods (a positive sign would indicate slowing down) and pupil size (a negative sign would indicate contraction).

Figures 33c-33f present the second-by-second normalized pupil size along several video durations. On visual inspection, there appears to be a large between-participant variability in the initial pupil size (1 second since video onset), compared to the same participant's average pupil size. At any rate, the initial response (contraction) is very rapid, as pupil size has stabilized at 2 seconds following video onset.

On average, linear regression slope was negative for all durations up to 16 seconds (suggestive of pupil contraction) and positive for longer durations, but in all cases, Bayesian one-sample t-test was supportive of the null hypothesis (mean slope being no

different than zero, $BF_{10}<1/3$), except for durations of 2 seconds (where Bayes factor indicated anecdotal evidence for the null, but borderline on the cut-off point of <1/3: $BF_{10}=0.355$) and 32 seconds (where the Bayes factor indicated strong evidence for a positive slope, $BF_{10}=16.452$). Repeated-measures ANOVA results consistently indicated evidence for the null hypothesis: no effect of time since video onset on second-by-second pupil size. This evidence for the null was only anecdotal for 2-second videos, and at least moderate for the rest. As indicated before, visual inspection on the evolution of pupil size (figures 32c-32f) hints that the pupil is comparatively dilated at video onset (1 second since) compared with later time points, but its size has rapidly reduced at 2 seconds since onset and does not contract further.

Thus, we did not find evidence for significant pupil size change throughout video presentation for the overall sample: however, we still wondered whether this could happen in a subset of trials, and whether there could be a potential association between heart rate and pupil size progression, specifically in relation to arousal-driven autonomic changes indicted in the heartbeat progression data above. To investigate this potential association, we calculated the slope for pupil size progression throughout each individual trial, as we had before for cardiac period progression, and normalized both within participant and video duration. Subsequently we ran a Bayesian ANCOVA on normalized pupil size slope, with normalized cardiac period slope and video duration as covariates and participant's ID as random factor. The likeliest model for prediction of pupil size slope included only cardiac period slope (besides ID), although its advantage over the second best model (the null model with only ID) was still anecdotal: $BF_{10}=2.770$. A Bayesian linear mixed-effects model on the effect of video duration and cardiac period slope (with ID as random-effect grouping variable) on pupil size slope rendered a B coefficient of -0.065 (95% credible intervals -0.121 - -0.009); the negative sign indicates that the association is in the direction predicted by our hypothesis: trials with a steeper slowing down of heart rate also have a more pronounced pupil contraction (figure 33g).

CHAPTER 14: DOPAMINERGIC ACTIVITY INDEXED BY SPONTANEOUS BLINKING AND DURATION ESTIMATION

Converging evidence points to a role of dopaminergic oscillators in time perception in intervals shorter and around 1 second duration. A study (Terhune et al., 2016) *took advantage of the interaction between dopamine activity and spontaneous blinking to test whether the latter could predict transient distortions in duration perception, and reported a lengthening of the perceived duration of short intervals if preceded by a blink within the last ~1 second. We investigated the same potential association in our data, but failed to find any relationship. We propose that, in the naturalistic conditions of our experimental set up (complex video scenes), the influence of dopaminergic fluctuations may be negligible in comparison with perceptual change driven by video content.*

As outlined in the Introduction to this study (Part V, Chapter 10), there is converging evidence to suggest that the ascending dopaminergic nigrostriatal pathway plays a critical role in both sub-second and supra-second time perception (Jennifer T Coull et al., 2011; Matell & Meck, 2004; Mauk & Buonomano, 2004). This has been indicated by selective lesion studies in animals (Dallal & Meck, 1993), pharmacological studies (T.H. Rammsayer, 1999), genetic (Wiener, Lee, Lohoff, & Costlett, 2014), neuroimaging (Melissa J. Allman & Warren H. Meck, 2012) and neuropsychological studies on both healthy volunteers and patients with neuropsychiatric disorders such as Parkinson's disease and schizophrenia (Melissa J. Allman & Warren H. Meck, 2012). Specifically, increased dopaminergic activity has been related to a lengthening of subjective time,

whereas dopamine depletion causes underestimation of durations (Jennifer T Coull et al., 2011). This has been explained by modulation by dopamine-based cortico-striatal circuits of the frequency and phase of cortical neural oscillators that are employed to track interval duration (Jennifer T Coull et al., 2011; Matell & Meck, 2004).

In light of these findings, Terhune and colleagues (Terhune et al., 2016) enquired whether transient fluctuations in dopaminergic activity could drive changes in duration judgments in healthy individuals. They employed spontaneous blinking as a proxy marker for dopaminergic activity, and found a bias in human responses towards reporting stimulus intervals (both sub-second auditory and visual; 300 - 700 milliseconds and supra-second visual stimuli; 1400 - 2600 milliseconds) as longer when the participant had blinked in the previous trial. In their experiment they employed simple auditory (white noise bursts) and visual stimuli (circles on a screen). Given that we have information about blinking thanks to eye-tracking recording, we decided to investigate whether the same association between blinking and duration estimation was present in our own data.

To do so, we categorized trials in function of whether there had been a blink within the 500 ms, as well as within the 2000 ms immediately leading to the onset of a given trial: thus we defined two binary variables based on those two time ranges, which were similar to those employed by Terhune and colleagues (they ascertained the presence of a blink in the previous trial, which had a sub-second or supra-second duration around the above mentioned values). Blinks were identified in the eye-tracking recording and it was the 'onset' of a blink within the two pre-trial time ranges that determined the binary classification. For brevity we will refer to these binary variables as B500 and B2000, respectively. In our experiment, trials started immediately following completion of the behavioural report for the previous trial. Consequently, the 500/2000 ms pre-onset periods correspond to the response phase of the previous trial: since it was not during video presentation, it is likely that these blinks were truly 'spontaneous', although we

cannot discard the possible influence of other effects, such as eye-fatigue from the previous trial, on these results.

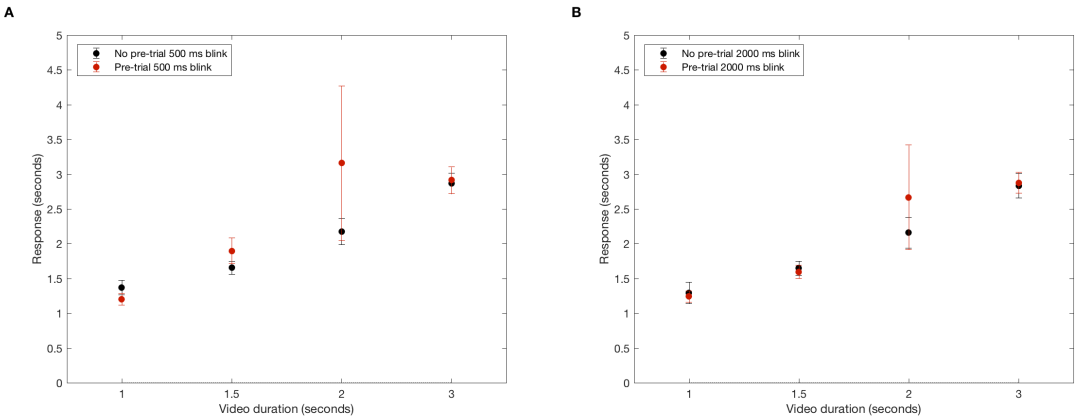


Figure 34. Duration estimation: Pre-trial blinking and behavioural results. Average response (seconds) by video duration according to presence or absence of a pre-trial blink, in the 500 ms (**34a**) or 2000 ms (**34b**) leading to video onset. The error bars represent between-participant standard error.

We ran a Bayesian ANCOVA on the whole dataset, with normalized response as dependent variable, B500 as fixed factor, video duration as covariate and participant's ID as random factor. We repeated the same analysis but using B2000 as fixed factor instead. In both cases, the null model was the most explanatory, followed by the model containing B500 or B2000 only: the null model outperformed these two by a factor of 12.82 and 24.39, respectively, indicating strong evidence in favour of the null compared to the second-best model in each case. Considering all possible models, there was strong ($BF_{inclusion}$=0.052) and very strong ($BF_{inclusion}$=0.028) evidence against inclusion of B500 and B2000 as predictor of response, respectively.

However, the presence of a blink during the 500 or 2000 ms leading to video onset would have little relationship with transient dopamine activity at the end of long video presentations - up to 64 seconds after the considered period. The modulatory activity

of dopaminergic pulses in terms of synchronizing cortical oscillators may be highly variable during the course of long videos, and exploration of transient fluctuations would be meaningless in those conditions.

Therefore, we restricted the analysis to durations up to 3 seconds - comprising 1248 trials. This range of durations also more closely corresponds to the supra-second range of values examined in (Terhune et al., 2016). Responses for those durations according to presence or absence of a pre-trial blink are presented in Figure 34. Nevertheless, after repeating the Bayesian RM ANOVAs in those conditions, results were very similar to those for the entire sample: the null model had the most evidence in both cases, outperforming the second-best model (which included only B500 or B2000, in each case) by a factor of 11.63 and 13.69, respectively. Overall, evidence for B500 or B2000 as predictor of normalized response was strongly against both ($BF_{inclusion}$=0.058 and 0.049, respectively). Likewise, repeating the analyses for each duration separately rendered moderate support for the null hypothesis (lack of effect of pre-trial blink on duration estimate, with BF_{10}<1/3 and >1/10) in all cases, except for B500 in 1-second videos, where the support for the null was anecdotal (BF_{10}=0.435).

In summary, in our own data we did not find any association between pre-trial blinking and subjective duration estimation, contrary to what would be expected given the proposed relationship between spontaneous blinking, dopamine activity and duration perception, and previous results (Terhune et al., 2016).

CHAPTER 15: DISCUSSION

Our analyses about trial-by-trial variation in duration estimates for naturalistic visual scenes support change in perceptual content (mainly visual variability over time and across the visual field) as a basis for human duration perception, while, in the conditions of our experiment, autonomic and dopaminergic activity indexed by heart rate, pupil size and blinking played no detectable part. We did observe an apparent parasympathetic response at the beginning of a new video, but the lack of association with duration estimates suggested an unspecific reaction to a novel stimulus. We surmise that, while in simplified lab conditions the influence of interoceptive content and neural oscillators may be greater, it is not the case in waking everyday conditions that constitute a major part of our conscious experience.

The mechanisms that give rise to our subjective perception of time are remain unclear. A fundamental role of autonomic processes (embodied time perception, usually conceptualized within the pacemaker-accumulator framework) (Treisman et al., 1990; Wittmann et al., 2010) or of dopamine-based neural oscillators operating at different timescales have often been proposed previously (Jennifer T Coull et al., 2011; Matell & Meck, 2004). However, contrary to such positions, most studies on the effect of autonomic processes have reported negative results (Bell & Provins, 1963; Schaefer & Gilliland, 1938), or else findings of doubtful interpretation (G. R. Hawkes, Joy, & Evans, 1961; T. Jamin et al., 2004; Lambourne, 2011; J. R. Tinklenberg, W. T. Roth, & B. S. Kopell, 1976; Vercruyssen et al., 1989) - given the relationship between autonomic processes, arousal and generic task performance (Dormal et al., 2017; Schwarz et al., 2013). On the other hand, studies on the role of neural oscillators suggest a role in the timing of short intervals (less or around 1 second) involved in motor activities (Merchant et al., 2014;

Merchant, Perez, Zarco, & Gamez, 2013), but its relevance for other instances of duration perception is less clear.

Importantly, most studies on duration perception employ simple, unchanging visual or auditory stimuli. This experimental setting obviates a long-standing intuition about subjective duration: that it is strongly affected by changes in perceptual content. A recent study (Roseboom et al., 2017) presented a model that produced duration estimates for natural video scenes (1-64 seconds long) on the basis of the amount of change in perceptual content, estimated through stimulus-driven changes in activation of an artificial perceptual classification network. Duration estimates produced by the model were able to reproduce estimates provided by human participants for the exact same videos, including being strongly affected by where in the scene content was coming from (based on human gaze) and being biased by the content of the scene, as also reported in the present study (Part V Chapter 12). To achieve this outcome, the model made no use of autonomic or oscillatory neural processes. The results reported in the present study further support the idea that repetitive autonomic or oscillatory neural pacemakers are not necessary to produce human subject duration estimation by suggesting that, at least when dynamic perceptual input is available (as in naturalistic scenes), humans do not make relevant use of such information in generating duration estimates. Specifically, we detected no effect of accumulated autonomic signals (cardiac activity and pupil size) nor blinks (indirectly indicating fluctuations in dopaminergic activity) on duration estimates in our data.

We did find evidence for an initial autonomic response at the beginning of a new video presentation, consisting of a rapid pupil contraction (occurring during the first 2 seconds since video onset) and a progressive cardiac deceleration during the first 5-10 seconds of video presentation. These two phenomena exhibited a highly variable intensity between trials and participants, but correlated with each other, suggesting a common underlying process, compatible with parasympathetic activation. Regarding cardiac deceleration, a similar phenomenon (albeit reportedly extending up to the end of the

presented intervals, lasting from 8 to 20 seconds) has been described as an indication of an accumulating pattern of autonomic activity working as a clock-type system, in line with an interoception-based pacemaker-accumulator model (Meissner & Wittman, 2011); in that study, cardiac deceleration was more pronounced in good performers. In our case, heart rate reduction and pupil contraction happen only at the beginning of video presentation, although for short videos this can comprise most or all of their duration: we therefore suppose that they reflect an initial arousal response in relation with a new video/task (Bradley, 2009), rather than processes specifically related to time perception. In relation to this point, visual inspection of our data seems to suggest a better accuracy in duration estimates for those trials or participants wherein such arousing response is more pronounced, but only for short videos -i.e. those where the initial response may contribute significantly to the amount of attention and engagement throughout the course of the entire video. Statistical analysis, however, did not reveal any association with accuracy. In fact, the vast majority of our analyses regarding the potential interaction of heart rate and both absolute duration estimation and duration estimation accuracy revealed evidence in favour of the null hypothesis that heart rate did not relate to duration estimation in our data. Even if such a potential association was present in our data, the simplest explanation would imply a generic association between arousal or attention and performance in a cognitive task. Our data does not support any association specifically related to the mechanism of time perception.

Regarding the lack of any relationship between blinking and duration perception, it is possible that the recorded blinks were not all truly spontaneous – our experiment was not explicitly designed to capture this. Blinks in our task may sometimes have been due to eye fatigue, or other similar factors, due to the long durations of trials and dynamic visual nature of the trials. Consequently, the association between blinks and with striatal dopaminergic activity would be less direct (though a similar criticism is also largely available for the original study) (Terhune et al., 2016). However, the complete lack of any statistical association between pre-trial blink and duration estimation suggests that

any such modulation had very little effect in the naturalistic conditions of our experiment.

Other than the difference between stimuli (our experiment used videos of natural scenes over a broader range of durations while the experiment by Terhune and colleagues (Terhune et al., 2016) used only basic auditory noise bursts or visual flashes), another difference between experiments was the task. In our experiment we required participants to directly estimate the observed duration in seconds. Terhune and colleagues (2016) used a temporal bisection task wherein participants reported whether the observed duration was closer to long or short anchor durations on which they were previously trained. While a large literature has linked striatal dopamine to duration estimation and time perception generally (M. J. Allman & W. H. Meck, 2012; Jennifer T Coull et al., 2011; Jones, Malone, Dirnberger, Edwards, & Jahanshahi, 2008; Matell & Meck, 2004; Mauk & Buonomano, 2004; T. H. Rammsayer, 1999; Wiener, Lee, Lohoff, & Coslett, 2014) , it is also well known that striatal dopamine is critically involved in the biasing of decisions in many other dimensions, especially for goal-directed or reward-driven decisions (K. Friston et al., 2014; Lepora & Gurney, 2012; Lo & Wang, 2006; Nagano-Saito et al., 2012). The absence of evidence for an effect of pre-trial blinks in our data may not reflect a failure to replicate the previously reported influence of spontaneous blinks on time perception, but instead that our direct estimation task is not susceptible to the kind of bias in report that is related to dopaminergic activity.

In summary, our results (together with results reported in (Roseboom et al., 2017) for the same dataset) strongly support the hypothesis that human subjective time perception (on the scale of seconds) is mainly driven by perceptual change, of which visual phenomenology represents a major part in naturalistic conditions -in other words, our results support a key role of visual variability throughout time and across the visual field in determining a high-level trait of conscious experience such as perception of the passage of time. It is likely than in other conditions than naturalistic wakeful experience (such as simplified laboratory conditions, resting state with closed eyes, etc)

interoceptive information plays a larger part in multi-modal perceptual content. Nevertheless, under the circumstances of our experiment, there is no evidence for a meaningful contribution of underlying autonomic or neural processes to duration estimation.

Bibliography

Addams, R. (1834). LI. An account of a peculiar optical phænomenon seen after having looked at a moving body. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 5*(29), 373-374. doi:10.1080/14786443408648481

Adler, W. T., & Ma, W. J. (2016). The computations underlying human confidence reports are probabilistic, but not Bayesian. *bioRxiv*.

Akrami, A., Kopec, C. D., Diamond, M. E., & Brody, C. D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature, 554*(7692), 368-372. doi:10.1038/nature25510

Alais, D., Ho, T., & Han, S. e. (2017). A Matched Comparison Across Three Different Sensory Pairs of Cross-Modal Temporal Recalibration From Sustained and Transient Adaptation. *i-Perception*, 1-18. doi:10.1177/2041669517718697

Alais, D., Lelung, J., & Burg, E. V. d. (2017). Linear Summation of Repulsive and Attractive Serial Dependencies: Orientation and Motion Dependencies Sum in Motion Perception. *The Journal of Neuroscience, 37*(16), 4381-4390.

Albretcht, A. R., & Scholl, B. J. (2010). Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychol Sci, 21*(4), 560-567. doi:10.1177/0956797610363543

Alexander, R. G., Schmidt, J., & Zelinsky, G. J. (2014). Are summary statistics enough? Evidence for the importance of shape in guiding visual search. *Visual Cognition, 22*(3-4), 595–609. doi:10.1080/13506285.2014.890989

Allman, M. J., & Meck, W. H. (2012). Pathophysiological Distortions in Time Perception and Timed Performance. *Brain, 135*(3), 656-677. doi:10.1093/brain/awr210

Allman, M. J., & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain, 135*(Pt 3), 656-677. doi:10.1093/brain/awr210

Almeida, J., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2008). Unconscious processing dissociates along categorical lines. *Proceedings of the National Academy of Sciences, 105*(39), 15214-15218. doi:10.1073/pnas.0805867105

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn Sci, 15*(3), 122-131. doi:10.1016/j.tics.2011.01.003

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol Sci, 15*(2), 106-111. doi:10.1111/j.0963-7214.2004.01502006.x

Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. *Psychol Sci, 19*(4), 392-398.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, 106*(18), 7345-7350. doi:10.1073/pnas.0808981106

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the Environment in Memory. *Psychol Sci, 2*(6), 396-408. doi:10.1111/j.1467-9280.1991.tb00174.x

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog Brain Res, 154*, 93-120. doi:10.1016/s0079-6123(06)54005-1

Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J. M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *J Neurosci, 22*(19), 8633-8646.

Anobile, G., Castaldi, E., Turi, M., Tinelli, F., & Burr, D. C. (2016). Numerosity but not texture-density discrimination correlates with math ability in children. *Dev Psychol, 52*(8), 1206-1216. doi:10.1037/dev0000155

Anobile, G., Cicchini, G. M., & Burr, D. C. (2013). Separate Mechanisms for Perception of Numerosity and Density. *Psychol Sci, 25*(1), 265-270. doi:10.1177/0956797613501520

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number As a Primary Perceptual Attribute: A Review. *Perception, 45*(1-2), 5-31. doi:10.1177/0301006615602599

Anobile, G., Turi, M., Cicchini, G. M., & Burr, D. C. (2015). Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *Journal of Vision, 15*(5), 4, 1-12.

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychol Sci, 12*(2), 157-162.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics, 70*(7), 1325-1326. doi:10.3758/pp.70.7.1325

Aristotle. (350 BCE). “De Somnis” (in Parva Naturalia, translated by Beare JI, 1931). In.

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proc Biol Sci, 281*(1797). doi:10.1098/rspb.2014.1791

Ashourian, P., & Loewenstein, Y. (2011). Bayesian Inference Underlies the Contraction Bias in Delayed Comparison Tasks. *PLOS One, 6*(5), e19551. doi:10.1371/journal.pone.0019551

Attarha, M., & Moore, C. M. (2015a). The capacity limitations of orientation summary statistics. *Atten Percept Psychophys, 77*(4), 1116-1131. doi:10.3758/s13414-015-0870-0

Attarha, M., & Moore, C. M. (2015b). The perceptual processing capacity of summary statistics between and within feature dimensions. *Journal of Vision, 15*(4), 9, 1-17.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol Rev, 61*(3), 183-193.

Au, R. K. C., Ono, F., & Watanabe, K. (2012). Time Dilation Induced by Object Motion is Based on Spatiotopic but not Retinotopic Positions. *Frontiers in Psychology, 3*(58). doi:10.3389/fpsyg.2012.00058

Bacon, F. (1854). *Bacon's Novum Organum*: T. And G. Shrimpton.

Bahrami, B., Carmel, D., Walsh, V., Rees, G., & Lavie, N. (2008). Spatial attention can modulate unconscious orientation processing. *Perception, 37*(10), 1520-1528. doi:10.1068/p5999

Balas, B., Nakano, L., & Rosenholtz, R. (2010). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12), 1301-1318. doi:10.1167/9.12.13

Balas, B., & Sinha, P. (2007). “Filling-in” colour in natural scenes. *Visual Cognition, 15*(7), 765-778. doi:10.1080/13506280701295453

Barlow, H. B. (2012). Possible Principles Underlying the Transformations of Sensory Messages. In *Sensory Communication*: The MIT Press.

Bauer, B. (2009). Does Stevens’s Power Law For Brightness Extend to Perceptual Brightness Averaging? *The Psychological Record, 59*(2).

Baumann, O., Endestad, T., Magnussen, S., & Greenlee, M. W. (2008). Delayed discrimination of spatial frequency for gratings of different orientation: behavioral and fMRI evidence for low-level perceptual memory stores in early visual cortex. *Exp Brain Res, 188*(3), 363-369. doi:10.1007/s00221-008-1366-0

Beach, L. R., & Scopp, T. S. (1967). *Intuitive statistical inferences about variances.*, Seattle.

Beach, L. R., & Swenson, R. G. (1966). Intuitive estimation of means. *Psychonomic Science, 5*(4), 161-162. doi:10.3758/bf03328331

Bell, C. R., & Provins, K. A. (1963). Relation between physiological responses to environmental heat and time judgments. *Journal of Experimental Psychology, 66*(6), 572-579.

Bentley, N. M., & Salinas, E. (2009). Neural Coding of Spatial Representations A2 - Squire, Larry R. In *Encyclopedia of Neuroscience* (pp. 117-122). Oxford: Academic Press.

Blake, R., Tadin, D., Sobel, K. V., Raissian, T. A., & Chong, S. C. (2006). Strength of early visual adaptation depends on visual awareness. *PNAS, 103*(12), 4783-4788. doi:10.1073 pnas.0509634103

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *bioRxiv*.

Block, R. A. (1982). Temporal judgments and contextual change. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(6), 530-544. doi:10.1037/0278-7393.8.6.530

Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 656-665. doi:10.1037/0278-7393.4.6.656

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nat Neurosci, 2*, 947. doi:10.1038/14731

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology, 46*(1), 1-11. doi:10.1111/j.1469-8986.2008.00702.x.

Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision, 15*(15), 6, 1-4.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *J Vis, 11*(5), 4. doi:10.1167/11.5.4

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We See More Than We Can Report: "Cost Free" Color Phenomenality Outside Focal Attention. . *Psychol Sci, 25*(7), 1394-1403. doi:10.1177/0956797614532656

brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics, 57*(1), 105-116. doi:10.3758/BF03211853

Burr, D., & Morrone, C. (2006). Perception: transient disruptions to neural space-time. *Curr Biol, 16*(19), R847-849. doi:10.1016/j.cub.2006.08.075

Burr, D., & Ross, J. (2008). A Visual Sense of Number. *Current Biology, 18*, 425-428.

Busemeyer, J. R., & Myung, I. J. (1987). Resource allocation decision making in an uncertain environment. *Acta Psychologica, 66*(1), 1-19. doi:https://doi.org/10.1016/0001-6918(87)90015-1

Campbell, F. W., & Maffei, L. (1971). The tilt after-effect: a fresh look. *Vision Res, 11*, 833-840.

Cant, J. S., & Xu, Y. (2012). Object Ensemble Processing in Human Anterior-Medial Ventral Visual Cortex. *The Journal of Neuroscience, 32*(22), 7685-7700. doi:10.1523/JNEUROSCI.3325-11.2012

Carandini, M., & Heeger, D. J. (2013). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51-62. doi:10.1038/nrn3136

Chalk, M., Marre, O., & Tkacik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. *Proc Natl Acad Sci U S A, 115*(1), 186-191. doi:10.1073/pnas.1711114115

Chang, A. Y., Schwartzman, D. J., VanRullen, R., Kanai, R., & Seth, A. K. (2017). Visual Perceptual Echo Reflects Learning of Regularities in Rapid Luminance Sequences. *J Neurosci, 37*(35), 8486-8497. doi:10.1523/jneurosci.3714-16.2017

Chen, C., Chen, X., Gao, M., Yang, Q., & Yan, H. (2015). Contextual influence on the tilt after-effect in foveal and para-foveal vision. *Neuroscience Bulletin, 31*(3), 307-316. doi:10.1007/s12264-014-1521-5

Chong, S., Joo, S., Emmanouil, T., & Treisman, A. (2008). Statistical processing: not so implausible after all. *Perception & Psychophysics, 70*(7), 1327-1334. doi:10.3758/PP.70.7.1327

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Res, 43*(4), 393-404.

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*(1), 1-13. doi:10.3758/bf03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: computing the average size in perceptual groups. *Vision Res, 45*, 891-900.

Church, R. M. (1984). Properties of the internal clock. In J. Gibbon & L. Allan (Eds.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and Time Perception* (pp. 52-77). New York: New York Academy of Sciences.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences, 111*(21), 7867-7872. doi:10.1073/pnas.1402785111

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision, 17*(14), 6, 1-9. doi:10.1167/17.14.6

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci, 36*(3), 181-204. doi:10.1017/s0140525x12000477

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends Cogn Sci, 16*(8), 411-417. doi:10.1016/j.tics.2012.06.013

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends Cogn Sci, 20*(5), 324-335. doi:10.1016/j.tics.2016.03.006

Collier, G., & Verplanck, W. S. (1958). Nonindependence of successive responses at the visual threshold as a function of interpolated stimuli. *Journal of Experimental Psychology, 55*(5), 429-437. doi:10.1037/h0047574

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating stable representations: serial dependence in vision. *PLOS One, 6*(1), e16701. doi:10.1371/journal.pone.0016701

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138*(2), 289-301.

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition, 20*(2), 211-231. doi:10.1080/13506285.2012.657261

Coull, J. T., Cheng, R.-K., & Meck, W. H. (2011). Neuroanatomical and Neurochemical Substrates of Timing. *Neuropsychopharmacology, 36*, 3-25.

Coull, J. T., Hwang, H. J., Leyton, M., & Dagher, A. (2012). Dopamine Precursor Depletion Impairs Timing in Healthy Volunteers by Attenuating Activity in Putamen and Supplementary Motor Area. *Journal of Neuroscience, 32*(47), 16704-16715. doi:10.1523/JNEUROSCI.1258-12.2012

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci, 24*(1), 87-114; discussion 114-185.

Craig, A. D. (2009). Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philosophical Transactions of the Royal Society of London, B Biological Science, 364*, 1933-1942.

Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nat Rev Neurosci, 10*(1), 59-70. doi:10.1038/nrn2555

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annu Rev Psychol, 66*, 115-142. doi:10.1146/annurev-psych-010814-015031

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space. *Neuron, 66*(6), 937-948. doi:10.1016/j.neuron.2010.05.018

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences, 108*(49), 19552-19557.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Res, 37*(22), 3181-3192.

Dallal, N. L., & Meck, W. H. (1993). Depletion of dopamine in the caudate nucleus but not destruction of vestibular inputs impairs short-interval timing in rats. *Abstracts-Society for Neuroscience, 19*, 1583.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Q J Exp Psychol (Hove), 62*(9), 1716-1722. doi:10.1080/17470210902811249

De Weerd, P., Gattass, R., Desimone, R., & Ungerleider, L. G. (1995). Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature, 377*(6551), 731-734. doi:10.1038/377731a0

DeCarlo, L. T. (2006). Sequential effects in successive ratio estimation. *Percept Psychophys, 68*(5), 861-871.

Di Luca, M., & Rhodes, D. (2016). Optimal Perceived Timing: Integrating Sensory Information with Dynamically Updated Expectations. *Sci Rep, 6*, 28563. doi:10.1038/srep28563

Ditchburn, R. W., & Ginsborg, B. L. (1952). Vision with a stabilized retinal image. *Nature, 170*(4314), 36-37.

Dormal, V., Heeren, A., Pesenti, M., & Maurage, P. (2017). Time Perception is not for the faint-hearted? Physiological arousal does not influence duration categorisation. *Cognitive Processing*. doi:10.1007/s10339-017-0852-3

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex. *Neuron, 28*, 287-298.

Drugowitsch, J., & Pouget, A. (2012). Probabilistic vs. non-probabilistic approaches to the neurobiology of perceptual decision-making. *Curr Opin Neurobiol, 22*(6), 963-969. doi:10.1016/j.conb.2012.07.007

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 149-169. doi:10.1037/0096-1523.21.1.149

Durgin, F. H. (1996). Visual aftereffect of texture density contigent on color of frame. *Percept Psychophys, 58*(2), 207-223.

Durgin, F. H. (2001). Texture contrast aftereffects are monocular; texture density aftereffects are binocular. *Vision Res, 41*(20), 2619-2630. doi:https://doi.org/10.1016/S0042-6989(01)00121-3

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology, 18*(18), R855-R856.

Durgin, F. H., & Hammer, J. T. (2001). Visual aftereffects of sequential perception: dynamic adaptation to changes in texture density and contrast. *Vision Res, 41*(20), 2607-2617.

Durgin, F. H., & Huk, A. C. (1997). Texture density aftereffects in the perception of artificial and natural textures. *Vision Res, 37*(23), 3273-3282.

Durgin, F. H., & Proffitt, D. R. (1996). Visual learning in the perception of texture: simple and contingent aftereffects of texture density. *Spat Vis, 9*(4), 423-474.

E., S. C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), 379-423. doi:doi:10.1002/j.1538-7305.1948.tb01338.x

Eagleman, D. M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Philos Trans R Soc Lond B Biol Sci, 364*(1525), 1841-1851. doi:10.1098/rstb.2009.0026

El-Shamayleh, Y., & Movshon, J. A. (2011). Neuronal responses to texture-defined form in macaque visual area V2. *J Neurosci, 31*(23), 8543-8555. doi:10.1523/jneurosci.5974-10.2011

Emmanouil, T., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics, 70*(6), 946-954.

Fairhall, A. L., Lewen, G. D., Blalek, W., & Steveninck, R. R. d. R. v. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature, 412*(6849), 787-792.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*, 292. doi:10.1038/nrn2258

Fan, J. E., Turk-Browne, N. B., & Taylor, J. A. (2016). Error-Driven Learning in Statistical Summary Perception. *Journal of Experimental Psychology. Human Perception and Performance., 42*(2), 266-280.

Fassihi, A., Akrami, A., Esmaeili, V., & Diamond, M. E. (2014). Tactile perception and working memory in rats and humans. *Proc Natl Acad Sci U S A, 111*(6), 2331-2336. doi:10.1073/pnas.1315171111

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front Hum Neurosci, 4*(215). doi:10.3389/fnhum.2010.00215

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nat Neurosci, 17*, 738-743. doi:10.1038/nn.3689

Fornaciai, M., Brannon, E. M., Woldorff, M. G., & Park, J. (2017). Numerosity processing in early visual cortex. *Neuroimage, 157*, 429-438.

Fornaciai, M., & Park, J. (2018). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychol Sci, 29*(3), 437-446. doi:10.1177/0956797617737385

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics, 70*(3), 456-464.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *J Exp Psychol Hum Percept Perform, 33*(5), 1003-1012. doi:10.1037/0096-1523.33.5.1003

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nat Neurosci, 14*(9), 1195-1201. doi:10.1038/nn.2889

Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 815-836. doi:10.1098/rstb.2005.1622

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philos Trans R Soc Lond B Biol Sci, 369*(1655). doi:10.1098/rstb.2013.0481

Friston, K. J., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology, 3*(151). doi:0.3389/fpsyg.2012.00151

Fritsche, M., Mostert, P., & Lange, F. P. d. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology, 27*, 1-6. doi:dx.doi.org/10.1016/j.cub.2017.01.006

Fründ, I., Wichmann, F. A., & Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision, 14*(7), 9, 1-16.

Gallistel, C. R., & Gibbon, J. (2000). Time, Rate and Conditioning. *Psychological Review, 107*(2), 289-344.

Gardelle, V. d., & Mammasian, P. (2015). Weighting Mean and Variability during Confidence Judgments. *PLOS One, 10*(3), e0120870. doi:10.1371/journal.pone.0120870

Gardelle, V. d., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, 108*(32), 13341-13346. doi:10.1073/pnas.1104517108

Geisler, W. S. (2008). Visual Perception and the Statistical Properties of Natural Scenes. *Annu Rev Psychol, 59*, 167-192. doi:10.1146/annurev.psych.58.110405.085632

Gheorghiu, E., Kingdom, F. A. A., & Petkov, N. (2014). Contextual modulation as de-texturizer. *Vision Res, 104*(2014), 12-23. doi:10.1016/j.visres.2014.08.013

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology, 12*, 453-467.

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proc Natl Acad Sci U S A, 97*(22), 12380-12384. doi:10.1073/pnas.97.22.12380

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn Sci, 17*(11), 585-593. doi:10.1016/j.tics.2013.09.001

Green, D., & Sweets, J. A. (1966). *Signal detection theory and ppsychophysics.*: New York: Krieger Publishing Company.

Greene, M. R., & Oliva, A. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychol Sci, 20*(4), 464-472.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci, 10*(1), 14-23. doi:10.1016/j.tics.2005.11.006

Grossberg, S., & Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Percept Psychophys, 43*(3), 241-277.

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual Differences in Ensemble Perception Reveal Multiple, Independent Levels of Ensemble Representation. *Journal of Experimental Psychology: General, 144*(2), 432-446. doi:0096-3445/15/$12.00

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision, 15*(4), 16, 11-11. doi:10.1167/15.4.16

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current biology : CB, 17*(17), 10.1016/j.cub.2007.1006.1039. doi:10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble procesing for sets of faces. *Journal of Experimental Psychology. Human Perception and Performance., 35*(3), 718-734. doi:10.1037/a0013899

Harris, J. P., & Calvert, J. E. (1985). The tilt after-effect: changes with stimulus size and eccentricity. *Spat Vis, 1*(2), 113-129.

Haun, A. M., Tononi, G., Koch, C., & Tsuchiya, N. (2017a). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness, 2017*(1), niw023-niw023. doi:10.1093/nc/niw023

Haun, A. M., Tononi, G., Koch, C., & Tsuchiya, N. (2017b). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness*, 1-4. doi:10.1093/nc/niw023

Hawkes, G. R., Joy, R. J., & Evans, W. O. (1961). Autonomic effects on estimates of time: evidence for a physiological correlate of temporal experience. *Rep US Army Med Res Lab*, 1-9.

Hawkes, G. R., Joy, R. J. T., & Evans, W. O. (1962). Autonomic effects on estimates of time: Evidence for a physiological correlate of temporal experience. *The Journal of Psychology: Interdisciplinary and Applied, 53*, 183-191.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*, 334-337.

He, S., Cohen, E. R., & Hu, X. (1998). Close correlation between activity in brain area MT/V5 and the perception of a visual motion aftereffect. *Current Biology, 8*, 1215-1218.

He, S., & MacLeod, D. I. A. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature, 411*, 473. doi:10.1038/35078072

https://www.nature.com/articles/35078072#supplementary-information

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Chapter 25 - Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements* (pp. 537-III). Oxford: Elsevier.

Herbst, S. K., Javadi, A. H., Meer, E. v. d., & Busch, N. A. (2013). How long depends on how fast -perceived flicker dilates subjective duration. *PLOS One, 8*(10), e76704.

Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N. W., McGraw, P. V., & Whitaker, D. (2012). Duration channels mediate human time perception. *Proceedings of the Royal Society B, 279*, 690-698. doi:10.1098/rspb.2011.1131

Hisakata, R., Nishida, S. y., & Johnston, A. (2016). An Adaptable Metric Shapes Perceptual Space. *Current Biology, 26*, 1911-1915.

Hofstatter, P. R. (1939). Uber die Schatzung von gruppeneigenschaften. *Zeitschrift fur Psychologie, 145*, 1-44.

Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology and Scientific Methods, 7*(17), 461-469.

Horr, N. K., & Di Luca, M. (2015). Filling the blanks in temporal intervals: the type of filling influences perceived duration and discrimination performance. *Front Psychol, 6*, 114. doi:10.3389/fpsyg.2015.00114

Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *J Vis, 10*(8), 11. doi:10.1167/10.8.11

Hsieh, P. J., & Tse, P. U. (2010). "Brain-reading" of perceived colors reveals a feature mixing mechanism underlying perceptual filling-in in cortical area V1. *Hum Brain Mapp, 31*(9), 1395-1407. doi:10.1002/hbm.20946

Huang, J., & Sekuler, R. (2014). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10*(2), 24.21-2427. doi:doi:10.1167/10.2.24

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci, 198*(1130), 1-59.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception, 43*(7), 663-676. doi:10.1068/p7719

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Atten Percept Psychophys, 75*(2), 278-286. doi:10.3758/s13414-012-0399-4

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093-1123. doi:10.1080/13506280444000661

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res, 49*(10), 1295-1306. doi:https://doi.org/10.1016/j.visres.2008.09.007

Jamin, T., Joulia, F., Fontanari, P., Giacomoni, M., Bonnon, M., Vidal, F., & Cremieux, J. (2004). Apnea-induced changes in time estimation and its relation to bradycardia. *Aviat Space Environ Med, 75*(10), 876-880.

Jamin, T., Joulia, F., Fontanari, P., Giacomoni, M., Bonnon, M., Vidal, F., & Crémieux, J. (2004). Apnea-Induced Changes in Time Estimation and its Relation to Bradycardia. *Aviation, Space, and Environmental Medicine, 75*(10), 876-880.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nat Neurosci, 9*(5), 690-697. doi:doi:10.1038/nn1691

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nat Neurosci, 13*(8), 1020-1026. doi:10.1038/nn.2590

John-Saaltink, E. S., Kok, P., Lau, H., & Lange, F. P. d. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience, 36*(23), 6186-6192.

Jones, C. R., Malone, T. J., Dirnberger, G., Edwards, M., & Jahanshahi, M. (2008). Basal ganglia, dopamine and temporal processing: performance on three timing tasks on and off medication in Parkinson's disease. *Brain Cogn, 68*(1), 30-41. doi:10.1016/j.bandc.2008.02.121

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*(5802), 91-97.

Kalm, K., & Norris, D. (2017). Visual recency bias is explained by a mixture model of short term memory. *bioRxiv*. doi:10.1101/228973

Kanai, R., Paffen, C. L., Hogendoorn, H., & Verstraten, F. A. (2006). Time dilation in dynamic visual display. *J Vis, 6*(12), 1421-1430. doi:10.1167/6.12.8

Kanai, R., Paffen, C. L. E., Hogendoorn, H., & Verstraten, F. A. J. (2011). Time dilation in dynamic visual display. *Journal of Vision, 6*, 1421-1471.

Kanai, R., & Verstraten, F. A. J. (2005). Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Res, 45*(2005), 3109-3116. doi:10.1016/j.visres.2005.05.014

Kanai, R., Wu, D. A., Verstraten, F. A., & Shimojo, S. (2006). Discrete color filling beyond luminance gaps along perceptual surfaces. *J Vis, 6*(12), 1380-1395. doi:10.1167/6.12.4

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the Misperception of Variability. *Journal of Experimental Psychology: General, 131*(2), 287-297. doi:10.1037//0096-3445.131.2.287

Karson, C. N. (1988). Physiology of normal and abnormal blinking. *Advances in Neurology, 49*, 25-37.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227-231. doi:10.1038/nature07200

Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A, 4*(12), 2395-2400. doi:10.1364/JOSAA.4.002395

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annu Rev Psychol, 55*, 271-304. doi:10.1146/annurev.psych.55.090902.142005

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Curr Opin Neurobiol, 13*(2), 150-158.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Curr Opin Neurobiol, 13*, 150-158.

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences, 21*(7), 493-497.

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends Cogn Sci, 21*(7), 493-497. doi:10.1016/j.tics.2017.04.011

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. *Journal of Vision, 10*(1), 8, 1-13. doi:10.1167/10.1.8

Kohn, A. (2007). Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *J Neurophysiol, 97*, 3155-3164. doi:10.1152/jn.00086.2007

Kok, R., Taubert, J., Burg, E. V. d., Rhodes, G., & Alais, D. (2016). Face familiarity promotes stable identity recognition: exploring face perception using serial dependence. *Royal Society Open Science*. doi:https://dx.doi.org/10.6084/m9.

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nat Rev Neurosci, 7*(3), 220-231. doi:10.1038/nrn1869

Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*(6971), 244-247. doi:10.1038/nature02169

Lambourne, K. (2011). The effects of acute exercise on temporal generalization. *The Quarterly Journal of Experimental Psychology, 65*(3), 526-540.

Lathrop, R. G. (1967). Perceived variability. *Journal of Experimental Psychology, 73*(4, Pt.1), 498-502. doi:10.1037/h0024344

Lediett, V., & Tong, J. E. (1972). Effects of heart-rate increase on temporal discrimination and time judgments by two groups of delinquents. *Perceptual and Motor Skills, 34*, 759-764.

Lee, M. D., & Wagenmakers, E.-J. (2013). Bayesian cognitive modeling: A practical course. . *Cambridge University Press*.

Lepora, N. F., & Gurney, K. N. (2012). The basal ganglia optimize decision making over general perceptual hypotheses. *Neural Comput, 24*(11), 2924-2945. doi:10.1162/NECO_a_00360

Lernia, D. D., Serino, S., Pezzulo, G., Pedroli, E., Cipresso, P., & Riva, G. (2018). Feel the time. Time perception as a function of interoceptive processing. *Front Hum Neurosci*. doi:10.3389/fnhum.2018.00074

Li, F. F., VanRullen, R., Koch, K., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, 99*(14), 9596-9601.

Li, V., Herce Castanon, S., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLoS Comput Biol, 13*(8), e1005723. doi:10.1371/journal.pcbi.1005723

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial Dependence in the Perception of Faces. *Current Biology, 24*(21), 2569-2574. doi:10.1016/j.cub.2014.09.025

Liberman, A., Manassi, M., & Whitney, D. (2018). Serial dependence promotes the stability of perceived emotional expression depending on face similarity. *Atten Percept Psychophys*. doi:10.3758/s13414-018-1533-8

Lin, Z., & He, S. (2009). Seeing the invisible: the scope and limits of unconscious processing in binocular rivalry. *Prog Neurobiol, 87*(4), 195-211. doi:10.1016/j.pneurobio.2008.09.002

Linares, D., & Gorea, A. (2015). Temporal frequency of events rather than speed dilates perceived duration of moving objects. *Scientific Reports, 5*(8825), 1-9.

Liu, W., Zhang, Z.-J., Zhao, Y.-J., Liu, Z.-F., & Li, B.-C. (2013). Effects of Awareness on Numerosity Adaptation. *PLOS One, 8*(10), e77556. doi:10.1371/journal.pone.0077556

Lo, C. C., & Wang, X. J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci, 9*(7), 956-963. doi:10.1038/nn1722

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *J Exp Psychol Hum Percept Perform, 9*(3), 461-473.

Lu, Z. L., Williamson, S. J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science, 258*(5088), 1668-1670.

Luca, M. D., & Rhodes, D. (2016). Optimal Perceived Timing: Integrating Sensory Information with Dynamically Updated Expectations. *Sci Rep*. doi:10.1038/srep28563

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn Sci, 17*(8), 391-400. doi:10.1016/j.tics.2013.06.006

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends Cogn Sci, 16*(10), 511-518. doi:10.1016/j.tics.2012.08.010

Magnussen, S., & Kurtenbach, W. (1980). Linear summation of tilt illusion and tilt aftereffect. *Vision Res, 20*, 39-42.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22*(6), 657-672. doi:10.3758/bf03209251

Malmierca, M. S., Anderson, L. A., & Antunes, F. M. (2015). The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding. *Front Syst Neurosci, 9*, 19. doi:10.3389/fnsys.2015.00019

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience, 26*, 147-155. doi:10.1017/S0952523808080905

Manassi, M., Liberman, A., Chaney, W., & Whitney, D. (2017). The perceived stability of scenes: serial dependence in ensemble representations. *Sci Rep, 7*(1), 1971. doi:10.1038/s41598-017-02201-5

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-018-1454-5

Martinez-Conde, S., & Macknik, S. L. (2017). Unchanging visions: the effects and limitations of ocular stillness. *Philos Trans R Soc Lond B Biol Sci, 372*(1718). doi:10.1098/rstb.2016.0204

Maruya, K., Watanabe, H., & Watanabe, M. (2008). Adaptation to invisible motion results in low-level but not high-level aftereffects. *Journal of Vision, 8*(11), 7, 1-11. doi:10.1167/8.11.7

Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research, 21*, 139-170.

Mather, G., Verstraten, F., & Anstis, S. (1998). *The Motion Aftereffect: A Modern Perspective*: The MIT Press.

Matthey, L., Bays, P. M., & Dayan, P. (2015). A Probabilistic Palimpsest Model of Visual Short-term Memory. *PLoS Comput Biol, 11*(1), e1004003. doi:10.1371/journal.pcbi.1004003

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience, 27*, 307-340.

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision, 15*(4), 1-18. doi:10.1167/15.4.6

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: metric and categorical effects on ensemble perception of hue. *Journal of the Optics Society of America. A: Optics, Image Science and Vision, 31*(4), A93-102. doi:10.1364/JOSAA.31.000A93

May, K. A., & Zhaoping, L. (2016). Efficient Coding Theory Predicts a Tilt Aftereffect from Viewing Untilted Patterns. *Current Biology, 26*, 1571-1576. doi:10.1016/j.cub.2016.04.037

Meck, W. H. (2003). Introduction: persistence of time. In W. H. Meck (Ed.), *Functional and Neural Mechanisms of Interval Timing* (pp. xvii-xli). Boca Raton, FL: CRC Press.

Meck, W. H. (2006). Neuroanatomical localization of an internal clock: a functional link between mesolimbic, nigrostriatal, and mesocortical dopaminergic systems. *Brain Research, 1109*(1), 93-107.

Meissner, K., & Wittman, M. (2011). Body signals, cardiac awareness, and the perception of time. *Biological Psychology, 86*, 289-297. doi:10.1016/j.biopsycho.2011.01.001

Merchant, H., Bartolo, R., Perez, O., Mendez, J. C., Mendoza, G., Gamez, J., . . . Prado, L. (2014). Neurophysiology of timing in the hundreds of milliseconds: multiple layers of neuronal clocks in the medial premotor areas. *Adv Exp Med Biol, 829*, 143-154. doi:10.1007/978-1-4939-1782-2_8

Merchant, H., Perez, O., Zarco, W., & Gamez, J. (2013). Interval tuning in the primate medial premotor cortex as a general timing mechanism. *J Neurosci, 33*(21), 9082-9096. doi:10.1523/jneurosci.5513-12.2013

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron, 88*, 78-92. doi:10.1016/j.neuron.2015.09.039

Michael, E., Gardelle, V. d., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences, 111*(21), 7873-7878. doi:10.1073/pnas.1308674111

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion. *Cognitive Computation, 3*(1), 5-24. doi:10.1007/s12559-010-9074-z

Morgan, M. J., Chubb, C., & Solomon, J. A. (2014). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11), 9.1-9.8. doi:10.1167/8.11.9

Morgan, M. J., Mareschal, I., Chubb, C., & Solomon, J. A. (2012). Perceived pattern regularity computed as a summary statistic: implications for camouflage. *Proceedings of the Royal Society of London. Series B, Biological Sciences., 279*, 2754-2760. doi:10.1098/rspb.2011.2645

Morgan, M. J., Raphael, S., Tibber, M. S., & Dakin, S. C. (2014). A texture-processing model of the 'visual sense of number'. *Proc Biol Sci, 281*(1790). doi:10.1098/rspb.2014.1137

Mori, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 52*(2), 72-83. doi:10.1037/h0087282

Münsterberg, H. (1899). THE PHYSIOLOGICAL BASIS OF MENTAL LIFE. *Science, 9*(221), 442.

Myczek, K., & Simons, D. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70*(5), 772-788.

Nagano-Saito, A., Cisek, P., Perna, A. S., Shirdel, F. Z., Benkelfat, C., Leyton, M., & Dagher, A. (2012). From anticipation to action, the role of dopamine in perceptual

decision making: an fMRI-tyrosine depletion study. *J Neurophysiol, 108*(2), 501-512. doi:10.1152/jn.00592.2011

Nguyen, V. T., Breakspear, M., Hu, X., & Guo, C. C. (2016). The integration of the internal and external milieu in the insula during dynamic emotional experiences. *Neuroimage, 124*, 455-463.

Noë, A. (2002). Is the Visual World a Grand Illusion? *Journal of Consciousness Studies, 9*(5-6), 1-12.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav Brain Sci, 24*(5), 939-973; discussion 973-1031.

Odegaard, B., Chang, M. Y., Lau, H., & Cheung, S.-H. (2018). Inflation versus filling-in: why we feel we see more than we actually do in peripheral vision. *bioRxiv*.

Ogmen, H., & Herzog, M. H. (2010). The Geometry of Visual Perception: Retinotopic and Non-retinotopic Representations in the Human Visual System. *Proc IEEE Inst Electr Electron Eng, 98*(3), 479-492. doi:10.1109/jproc.2009.2039028

Okazawa, G., Tajima, S., & Komatsu, H. (2017). Gradual Development of Visual Texture-Selective Properties Between Macaque Areas V2 and V4. *Cereb Cortex, 27*(10), 4867-4880. doi:10.1093/cercor/bhw282

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Prog Brain Res, 155*, 23-36. doi:10.1016/s0079-6123(06)55002-2

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: effects of internal and external noise. *J Vis, 14*(11). doi:10.1167/14.11.5

Ornstein, R. E. (1969). *On the Experience of Time*: Harmondsworth.

Osato, E., Ogawa, N., & Takaoka, N. (1995). Relations among heart rate, immediate memory, and time estimation under two different instructions. *Perceptual and Motor Skills, 80*, 831-842.

Otten, M., Pinto, Y., Paffen, C. L. E., Seth, A. K., & Kanai, R. (2016). The Uniformity Illusion: Central Stimuli Can Determine Peripheral Perception. *Psychol Sci*, 1-13. doi:10.1177/0956797616672270

Papadimitriou, C., Ferdoash, A., & Snyder, L. H. (2015). Ghosts in the machine: memory interference from the previous trial. *J Neurophysiol, 113*(2), 567-577. doi:10.1152/jn.00402.2014

Parker, D. M. (1972). Contrast and size variables and the tilt after-effect. *Q J Exp Psychol, 24*(1), 1-7. doi:10.1080/14640747208400260

Parkes, L., Lund, J., Angelucci, A., Solomon, J., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci, 4*(7), 739-744.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci, 4*(7), 739-744. doi:10.1038/89532

Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct Effects of Brief and Prolonged Adaptation on Orientation Tuning in Primary Visual Cortex. *The Journal of Neuroscience, 33*(2), 532-543. doi:10.1523/JNEUROSCI.3345-12.2013

Payzan-LeNestour, E., Balleine, B. W., Berrada, T., & Pearson, J. (2016). Variance After-Effects Distort Risk Perception in Humans. *Current Biology, 26*, 1-5. doi:http://dx.doi.org/10.1016/j.cub.2016.04.023

Peterson, C. R., & Beach, L. R. (1967). Man as a Intuitive Statistician. *Psychological Bulletin, 68*(1), 29-46.

Petrov, Y., Popple, A. V., & McKee, S. P. (2007). Crowding and surround suppression: Not to be confused. *Journal of Vision, 7*(2), 12-12. doi:10.1167/7.2.12

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian Estimation as an Explanation for Range and Regression Effects: A Study on Human Path Integration. *The Journal of Neuroscience, 31*(47), 17220-17229.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends Cogn Sci, 19*(5), 285-293. doi:https://doi.org/10.1016/j.tics.2015.03.002

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547-555. doi:10.1016/j.neuron.2004.10.014

Podvalny, E., Yeagle, E., Megevand, P., Sarid, N., Harel, M., Chechik, G., . . . Malach, R. (2017). Invariant Temporal Dynamics Underlie Perceptual Stability in Human Visual Cortex. *Curr Biol, 27*(2), 155-165. doi:10.1016/j.cub.2016.11.024

Pollack, I. W., Ochberg, F. M., & Meyer, E. (1965). EFFECT OF DEPRIVATION OF SIGHT ON SUBJECTIVE TIME SENSE. *Psychosom Med, 27*, 71-79.

Pollatos, O., Yeldesbay, A., Pikovsky, A., & Rosenblum, M. (2014). How much time has passed? Ask your heart. *Frontiers in Neurorobotics*. doi:10.3389/fnbot.2014.00015

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium*: Lawrence Erlbaum.

Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition, 11*(1), 77-82. doi:10.3758/BF03197664

Poynter, W. D., & Homa, D. (1983). Duration judgment and the experience of change. *Perception & Psychophysics, 33*(6), 548-560. doi:10.3758/BF03202936

Preuschhof, C., Schubert, T., Villringer, A., & Heekeren, H. R. (2010). Prior Information biases stimulus representations during vibrotactile decision making. *J Cogn Neurosci, 22*(5), 875-887. doi:10.1162/jocn.2009.21260

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. http://www.palamedestoolbox.org.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat Vis, 3*(3), 179-197.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychol Sci, 26*(11), 1664-1680. doi:10.1177/0956797615595037

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., Lange, F. P. d., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nat Neurosci, 14*(12), 1513-1515. doi:10.1038/nn.2948

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J Neurophysiol, 107*, 1556–1563. doi:doi:10.1152/jn.00985.2011

Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *Q J Exp Psychol B, 52*(3), 273-286. doi:10.1080/713932708

Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *The Quarterly Journal of Experimental Psychology. B., 52*, 273-286.

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3*(1), 1-23.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79-87. doi:10.1038/4580

Raphael, S., & Morgan, M. J. (2016). The computation of relative numerosity, size and density. *Vision Res, 124*, 15-23. doi:10.1016/j.visres.2014.12.022

Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: the normative approach and its heuristic approximation. *PLoS Comput Biol, 8*(10), e1002731. doi:10.1371/journal.pcbi.1002731

Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society B, 278*, 1314-1322. doi:10.1098/rspb.2010.1737

Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *PNAS, 114*(2), 412-417. doi:www.pnas.org/cgi/doi/10.1073/pnas.1610706114

Romo, R., & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci, 4*(3), 203-218. doi:10.1038/nrn1058

Roseboom, W. (2017). Serial dependence across multisensory relative timing tasks. *PsyArxiv*. doi:10.17605/OSF.IO/6BKDA

Roseboom, W., Fountas, Z., Nikiforou, K., Bhowmik, D., Shanahan, M., & Seth, A. K. (2017). Time without clocks: Human time perception based on perceptual classification. *BiorXiv*. doi:http://dx.doi.org/10.1101/172387

Roseboom, W., Linares, D., & Nishida, S. y. (2015). Sensory adaptation for timing perception. *Proceedings of the Royal Society B, 282*, 20142833. doi:10.1098/rspb.2014.2833

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A Summary Statistic Representation in Peripheral Vision Explains Visual Search. *Journal of Vision, 12*(4), 1-29. doi:10.1167/12.4.14

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision, 7*(14), 16-16. doi:10.1167/7.14.16

Samaha, J., Switzky, M., & Postle, B. R. (2018). Confidence boosts serial dependence in orientation estimation. *bioRxiv*. doi:10.1101/369140

Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision, 14*(13), 13, 11-13. doi:10.1167/14.13.13

Schaefer, V. G., & Gilliland, A. R. (1938). The relation of time estimation to certain physiological changes. *Journal of Experimental Psychology, 23*, 545-552.

Schieting, S., & Spillmann, L. (1987). Flicker adaptation in the peripheral retina. *Vision Res, 27*(2), 277-284.

Schummers, J., Sharma, J., & Sur, M. (2005). Bottom-up and top-down dynamics in visual cortex. *Prog Brain Res, 149*, 65-81. doi:10.1016/s0079-6123(05)49006-8

Schwarz, M. A., Winkler, I., & Sedlmeier, P. (2013). The heart beat does not make us tick: The impacts of heart rate and arousal on time perception. *Attention, Perception & Psychophysics, 75*, 182-193. doi:10.3758/s13414-012-0387-8

Schwiedrzik, C. M., Ruff, C. C., Lazar, A., Leitner, F. C., Singer, W., & Melloni, L. (2014). Untangling perceptual memory: hysteresis and adaptation map into separate cortical networks. *Cereb Cortex, 24*(5), 1152-1164. doi:10.1093/cercor/bhs396

Senders, V. L., & Sowards, A. (1952). Analysis of Response Sequences in the Setting of a Psychophysical Experiment. *The American Journal of Psychology, 65*(3), 358-374.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *J Vis, 13*(8). doi:10.1167/13.8.16

Solovey, G., Graney, G. G., & Lau, H. (2015). A decisional account of subjective inflation of visual perception at the periphery. *Atten Percept Psychophys, 77*, 258-271. doi:10.3758/s13414-014-0769-1

Spence, M., Dux, P., & Arnold, D. (2016). Computations Underlying Confidence in Visual Perception. *Journal of Experimental Psychology: Human Perception and Performance, 42*(5), 671-682. doi:10.1037/xhp0000179

Spencer, J. (1961). Estimating averages. *Ergonomics, 4*, 317-328. doi:10.1080/00140136108930533

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt aftereffect. *Percept Psychophys, 62*(8), 1525-1533.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput, 31*(1), 137-149.

Stein, T., & Sterzer, P. (2011). High-level face shape adaptation depends on visual awareness: Evidence from continuous flash suppression. *Journal of Vision, 11*(8), 5,1-14. doi:10.1167/11.8.5

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics, 1*(5), 439-446. doi:10.3758/bf03207424

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology, 6*, 157:151-154. doi:10.3389/fpsyg.2015.00157

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11*(5), 13, 11-82.

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018a). The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence from Sensory Adaptation. *i-Perception, 9*(5), 1-12. doi:10.1177/2041669518800728

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018b). Serial dependence in the perception of visual variance. *Journal of Vision, 18*(7), 4-4. doi:10.1167/18.7.4

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Atten Percept Psychophys, 76*(7), 2071-2079. doi:10.3758/s13414-014-0690-7

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn Sci, 13*(9), 403-409. doi:10.1016/j.tics.2009.06.003

Summerfield, C., & Lange, F. P. d. (2014). Expectation in perceptual decision-making: neural and computational mechanisms. *Nature Reviews Neuroscience, 15*, 745-756. doi:10.1038/nrn3838

Sun, H.-C., Jr., C. L. B., & Kingdom, F. A. A. (2016). Simultaneous density contrast is bidirectional. *Journal of Vision, 16*(14), 4, 1-11.

Sun, H.-C., Kingdom, F. A. A., & Curtis L. Baker, J. (2017). Texture density adaptation can be bidirectional. *Journal of Vision, 17*(8), 9: 1-10. doi:10.1167/17.8.9

Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychol Sci, 25*(10), 1903-1913. doi:10.1177/0956797614544510

Symonds, R. M., Lee, W. W., Kohn, A., Schwartz, O., Witkowski, S., & Sussman, E. S. (2017). Distinguishing Neural Adaptation and Predictive Coding Hypotheses in Auditory Change Detection. *Brain Topogr, 30*(1), 136-148. doi:10.1007/s10548-016-0529-8

T. Buswell, G. (1935). *How people look at pictures: a study of the psychology and perception in art*.

Tanaka, H., & Ohzawa, I. (2009). Surround suppression of V1 neurons mediates orientation-based representation of high-order visual features. *J Neurophysiol, 101*(3), 1444-1462. doi:10.1152/jn.90749.2008

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu Rev Neurosci, 19*, 109-139. doi:10.1146/annurev.ne.19.030196.000545

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *J Vis, 11*(5), 5. doi:10.1167/11.5.5

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Sci Rep*. doi:10.1038/srep32239

Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics, 24*(4), 305-314. doi:10.3758/bf03204247

Terao, M., Watanabe, J., Yagi, A., & Nishida, S. (2008). Reduction of stimulus visibility compresses apparent time intervals. *Nat Neurosci, 11*(5), 541-542. doi:10.1038/nn.2111

Terhune, D. B., Sullivan, J. G., & Simola, J. M. (2016). Time dilates after spontaneous blinking. *Current Biology, 26*, R445-460.

Thielscher, A., Kolle, M., Neumann, H., Spitzer, M., & Gron, G. (2008). Texture segmentation in human perception: a combined modeling and fMRI study. *Neuroscience, 151*(3), 730-736. doi:10.1016/j.neuroscience.2007.11.040

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision, 12*(6), 8-8. doi:10.1167/12.6.8

Tinklenberg, J. R., Roth, W. T., & Kopell, B. S. (1976). Marijuana and ethanol: differential effects on time perception, heart rate, and subjective response. *Psychopharmacology (Berl), 49*(3), 275-279.

Tinklenberg, J. R., Roth, W. T., & Kopell, B. S. (1976). Marijuana and ethanol: Differential effects on time perception, heart rate, and subjective response. *Psychopharmacology, 53*, 183-191.

Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The Internal Clock: Evidence for a Temporal Oscillator Underlying Time Perception with Some Estimates of its Characteristic Frequency. *Perception, 19*(6), 705-743.

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Percept Psychophys, 66*(7), 1171-1189.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat Neurosci, 8*(8), 1096-1101. doi:10.1038/nn1500

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131. doi:10.1126/science.185.4157.1124

Tynan, P., & Sekular, R. (1975). Moving visual phantoms: a new contour completion effect. *Science, 188*(4191), 951-952.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*(3), 473-494. doi:10.1080/13506280444000409

van Tuijl, H. F., & Leeuwenberg, E. L. (1979). Neon color spreading and structural information measures. *Percept Psychophys, 25*(4), 269-284.

Vercruyssen, M., Hancock, P., & Mihaly, T. (1989). Time estimation performance before, during, and following physical activity. *Journal of Human Ergology, 18*(2), 169-179.

Visscher, K. M., Kahana, M. J., & Sekuler, R. (2009). Trial-to-trial carry-over of item- and relational-information in auditory short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*(1), 46-56. doi:10.1037/a0013412

Wade, N. J. (1998). Early Studies of Eye Dominances. *Laterality: Asymmetries of Body, Brain and Cognition, 3*(2), 97-108. doi:10.1080/713754296

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1-19. doi:https://doi.org/10.3758/s13423-017-1323-7

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual system averages speed information. *Vision Res, 32*(5), 931-941. doi:https://doi.org/10.1016/0042-6989(92)90036-I

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Res, 29*(1), 47-59. doi:https://doi.org/10.1016/0042-6989(89)90173-9

Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision, 11*(5), 10.1167/1111.1165.1163 1163. doi:10.1167/11.5.3

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nat Neurosci, 18*(10), 1509-1517. doi:10.1038/nn.4105

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain, 97*(4), 709-728.

Wenderoth, P., & Johnstone, S. (1988). The different mechanisms of the direct and indirect tilt illusions. *Vision Res, 28*(2), 301-312.

Whitney, D., & Levi, D. M. (2011). Visual Crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn Sci, 15*(4), 160-168. doi:10.1016/j.tics.2011.02.005

Wiener, M., Lee, Y. S., Lohoff, F. W., & Coslett, H. B. (2014). Individual differences in the morphometry and activation of time perception networks are influenced by dopamine genotype. *Neuroimage, 89*, 10-22. doi:10.1016/j.neuroimage.2013.11.019

Wiener, M., Lee, Y. S., Lohoff, F. W., & Costlett, H. B. (2014). Individual differences in the morphometry and activation of time perception networks are influenced by dopamine genotype. *Neuroimage, 89*, 10-22.

Wittmann, M., Simmons, A. N., Aron, J. L., & Paulus, M. P. (2010). Accumulation of neural activity in the posterior insula encodes the passage of time. *Neuropsychologia, 48*(10), 3110-3120. doi:10.1016/j.neuropsychologia.2010.06.023

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision, 15*(4), 11:11-13. doi:10.1167/15.4.11

Wu, D. A., Kanai, R., & Shimojo, S. (2004). Steady-state misbinding of colour and motion. *Nature, 429*(6989), 262. doi:10.1038/429262a

Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., & Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Comput, 18*(1), 80-118. doi:10.1162/089976606774841585

Xia, Y., Liberman, A., Leib, A. Y., & Whitney, D. (2015). Serial Dependence in the perception of attractiveness. *Journal of Vision, 15*(12), 1219. doi:10.1167/15.12.1219

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cogn Psychol, 24*(3), 295-340.

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature, 414*(6861), 302-305. doi:10.1038/35104551

Yu, A. J., & Cohen, J. D. (2008). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems, 21*, 1873-1880.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 123-162). Cambridge: Cambridge University Press.

Zakay, D., Tsal, Y., Moses, M., & Shahar, I. (1994). The role of segmentation in prospective and retrospective time estimation processes. *Memory & Cognition, 22*(3), 344-351. doi:10.3758/BF03200861

Zavitz, E., & Baker, C. L. (2013). Texture sparseness, but not local phase structure, impairs second-order segmentation. *Vision Res, 91*, 45-55. doi:10.1016/j.visres.2013.07.018

Zavitz, E., & Baker, C. L., Jr. (2014). Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues. *J Vis, 14*(4). doi:10.1167/14.4.14

Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychol Sci, 22*(9), 1212-1219. doi:10.1177/0956797611419304

Ziemba, C., & Simoncelli, E. (2015). Opposing effects of summary statistics on peripheral discrimination. *Journal of Vision, 15*(12), 770. doi:10.1167/15.12.770

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016). Selectivity and tolerance for visual texture in macaque V2. *Proc Natl Acad Sci U S A, 113*(22), E3140-3149. doi:10.1073/pnas.1510847113

Zohar Z. Bronfman, N. B., Hilla Jacobson, Marius Usher. (2014). We See More Than We Can Report: “Cost Free” Color Phenomenality Outside Focal Attention. *Psychological Science, 25*(7), 1394-1403. doi:10.1177/0956797614532656

Zorzi, M., Di Bono, M. G., & Fias, W. (2011). Distinct representations of numerical and non-numerical order in the human intraparietal sulcus revealed by multivariate pattern recognition. *Neuroimage, 56*(2), 674-680. doi:10.1016/j.neuroimage.2010.06.035

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness & Cognition, 27*, 246-253. doi:dx.doi.org/10.1016/j.concog.2014.05.012

Appendix

PUBLISHED OR ACCEPTED ARTICLES

Homocysteine and cognition: A systematic review of 111 studies

I contributed to this article as a side project during my PhD, by performing a meta-analysis for the available evidence about the relationship between homocysteine levels and cognition in general population and patients with neuropsychiatric disorders.

Neuroscience and Biobehavioral Reviews 69 (2016) 280–298

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Homocysteine and cognition: A systematic review of 111 studies

Esther Setién-Suero[a,b,c,*,1], Marta Suárez-Pinilla[d,1], Paula Suárez-Pinilla[a,b,c], Benedicto Crespo-Facorro[a,b,c], Rosa Ayesa-Arriola[a,b,c]

[a] University Hospital Marqués de Valdecilla, Department of Psychiatry, School of Medicine, University of Cantabria, Santander, Spain
[b] CIBERSAM, Biomedical Research Network on Mental Health Area, Madrid, Spain
[c] IDIVAL, Valdecilla Biomedical Research Institute, Santander, Spain
[d] Sackler Centre for Consciousness Science, Department of Informatics, University of Sussex, United Kingdom

http://dx.doi.org/10.1016/j.neubiorev.2016.08.014

Serial dependence in the perception of visual variance

The contents of this paper form a large part of Part II of this thesis.

Journal of Vision (2018) 18(7):4, 1–24

Serial dependence in the perception of visual variance

Marta Suárez-Pinilla
Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK

Anil K. Seth
Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK

Warrick Roseboom
Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK

The recent history of perceptual experience has been shown to influence subsequent perception. Classically, this dependence on perceptual history has been examined in sensory-adaptation paradigms, wherein prolonged exposure to a particular stimulus (e.g., a vertically oriented grating) produces changes in perception of subsequently presented stimuli (e.g., the tilt aftereffect). More recently, several studies have investigated the influence of shorter perceptual exposure with effects, referred to as *serial dependence*, being described for a variety of low- and high-level perceptual dimensions. In this study, we examined serial dependence in the processing of dispersion statistics, namely variance—a key descriptor of the environment and indicative of the precision and reliability of ensemble representations. We found two opposite serial dependences operating at different timescales, and likely originating at different processing levels: A positive, Bayesian-like bias was driven by the most recent exposures, dependent on feature-specific decision making and appearing only when high confidence was placed in that decision; and a longer lasting negative bias—akin to an adaptation aftereffect—becoming manifest as the positive bias declined. Both effects were independent of spatial presentation location and the similarity of other close traits, such as mean direction of the visual variance stimulus. These findings suggest that visual variance processing occurs in high-level areas but is also subject to a combination of multilevel mechanisms balancing perceptual stability and sensitivity, as with many different perceptual dimensions.

Introduction

Considerable evidence indicates that the human visual system is able to extract statistical information from sensory signals supporting the formation of summary or ensemble representations across a variety of dimensions, including low-level features such as orientation or size, as well as higher level (complex or abstract) traits such as facial expressions in a crowd (Alvarez, 2011; Alvarez & Oliva, 2009; Ariely, 2001; Chong & Treisman, 2003; Geisler, 2008; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Such information can be used to efficiently encode (Dahmen, Keating, Nodal, Schulz, & King, 2010; Fairhall, Lewen, Blalek, & de Ruyter van Steveninck, 2001) and interpret subsequent sensory inputs, and to make predictions about future events (Summerfield & de Lange, 2014; Summerfield & Egner, 2009).

Many forms of visual input can be summarized in terms of statistical moments such as central tendency (e.g., mean) and variance or dispersion (consider, for example, a random-dot kinematogram, which will have a mean and a variance in the distribution of dot motion). Most studies on ensemble processing have focused on central-tendency statistics (Albrecht & Scholl, 2010; Chong & Treisman, 2005; Corbett & Oriet, 2011; Haberman & Whitney, 2009; Im & Chong, 2014; Sweeny & Whitney, 2014; Wolfe, Kosovicheva, Leib, Wood, & Whitney, 2015), while variance computations have received less attention. However, variance is known to play a crucial role in visual experience, modulating perceptual grouping, ensemble averaging (Brady & Alvarez, 2015; de Gardelle & Mammasian, 2015; de Gardelle & Summerfield, 2011;

Citation: Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision, 18*(7):4, 1–24, https://doi.org/10.1167/18.7.4.

https://doi.org/10.1167/18.7.4 Received January 17, 2018; published July 2, 2018 ISSN 1534-7362

Maule & Franklin, 2015; Maule, Witzel, & Franklin, 2014; Zylberberg, Roelfsema, & Sigman, 2014), texture processing (Morgan, Chubb, & Solomon, 2014; Morgan, Mareschal, Chubb, & Solomon, 2012), and comparison between arrays (Fouriezos, Rubenfeld, & Capstick, 2008). Variance is also critical to perceptual prediction, since it provides a measure of the expected range of stimuli (Summerfield & de Lange, 2014) as well as the precision (reliability) of the sensory input (Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Meyniel, Sigman, & Mainen, 2015; Sato & Kording, 2014). As an indication of sensory reliability, variance also affects metacognitive judgments, although evidence is conflicting regarding the extent and direction of this effect (de Gardelle & Mammasian, 2015; Spence, Dux, & Arnold, 2016; Zylberberg et al., 2014).

Notably, most studies involving variance manipulations have examined its impact on perceptual decisions about other features rather than on the perception of variance itself. Comparatively few studies have investigated the mechanisms of variance processing directly. Those that do have addressed mainly three questions: What are the general properties of variance processing (speed, automaticity, attentional demands)? To what extent does variance computation rely on the processing of the individual elements of the ensemble? And does it operate as a specific trait of the ensemble or feature dimension over which it is computed, or rather as an abstract property? So far, these studies have employed heterogeneous designs and reached disparate conclusions. Concerning the general properties of variance processing, a study examining judgments of color diversity (Bronfman, Brezis, Jacobson, & Usher, 2014) suggested a rapid, preattentive mechanism. This is in agreement with another study which reported priming by visual variance, an effect that seems to occur rapidly, automatically, and without need of feature-based attention (Michael, de Gardelle, & Summerfield, 2014); however, this latter study did not examine variance perception directly, but only the priming effect of variance on mean judgments.

Regarding the second question, namely the reliance of variance computation on the processing of individual elements, available evidence (based on highly heterogeneous studies) is conflicting: One study on pattern regularity (positional variance) suggested a very inefficient computation of variance, underwritten by subsampling of a small fraction of the elements of the array (Morgan et al., 2012). By contrast—and surprisingly, given the finding of a rapid, preattentive mechanism—the aforementioned study on color diversity reported that variance processing required a conscious representation of the individual elements (Bronfman et al., 2014). In a similar manner, a study on facial emotions in a crowd determined that variance judgments along this dimension relied on high-level processing of individual faces (Haberman, Lee, & Whitney, 2015).

Finally, regarding the question whether variance, once computed over a certain feature dimension of a visual ensemble, retains its specificity or emerges as an abstract property, the previously reported study on priming suggested that the effect of (implicit) variance on mean perception was feature specific (Michael et al., 2014). In contrast, a study on direct variance perception found negative adaptation aftereffects which generalized across dimensions of visual variance, suggesting a high-level rather than a sensory origin for this effect, and therefore indicating that variance may operate as an independent cognitive property (Payzan-LeNestour, Balleine, Berrada, & Pearson, 2016). In summary, available evidence shows some dissension, but a picture starts to emerge: Variance computations would be relatively rapid, but appear to require high-level processing of the individual elements of the array; however, once computed, variance would become dissociated from the properties of the ensemble and of the perceptual dimension over which it was estimated, and operate as a high-level cognitive trait.

To clarify some these issues, here we examine variance processing as a distinct perceptual feature by investigating the influence of previous variance presentations on judgments about this dimension. It has long been known that perception is affected by previous input. Influences of past perceptual events on current perception fall generally into two different categories. Best known are adaptation aftereffects—repulsive (negative) biases in perception exerted after prolonged exposure to a certain stimulus magnitude—which have been described for many low- and high-level traits, including variance (Campbell & Maffei, 1971; Mather, Verstraten, & Anstis, 1998; Payzan-LeNestour et al., 2016; Roseboom, Linares, & Nishida, 2015), and which are classically employed as an experimental tool for investigating perceptual mechanisms (Kohn, 2007). The second category is observed in relation to shorter presentations, generally consisting of an attractive (positive) perceptual bias toward recent sensory history. These *serial dependences* have been found for several low- and high-level features (Cicchini, Anobile, & Burr, 2014; Fischer & Whitney, 2014; John-Saaltink, Kok, Lau, & de Lange, 2016; Liberman, Fischer, & Whitney, 2014; Xia, Liberman, Leib, & Whitney, 2015). It has been proposed that these two different effects contribute in opposite ways to the tuning of the balance between perceptual sensitivity and stability: While negative adaptation produces a normalization of neural representations in order to maximize sensitivity to changes around the most frequent stimulus intensity, serial dependence contributes to perceptual stability by smoothing out discrete

discontinuities as sensory noise (Fischer & Whitney, 2014).

Our study employs serial dependence in variance judgments as a way to track the dynamics and timescales of the processing of this statistic as a distinct perceptual feature. We conducted three experiments.

Experiment 1 investigated the existence, magnitude, and direction of serial dependence in visual variance perception, as well as its relationship with associated stimulus features such as ensemble mean and spatial location. In addition, we separately explored fovea and periphery, as the compression of visual information into summary statistics is particularly relevant to the much larger, poorly spatially resolved peripheral field (Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Rosenholtz et al., 2012; Ziemba & Simoncelli, 2015). Experiment 2 attempted to identify the specific levels of perceptual decision that give rise to serial dependence in variance: whether it is a bias in low-level perceptual, decision-making, or response processes. In Experiment 3 we investigated the relationship between the reported confidence in perceptual decisions and their influence in subsequent judgments along the same dimension, especially considering Bayesian accounts of confidence as a measure of the precision of neural representations (Meyniel et al., 2015).

Overall, our results indicate that judgments of visual variance are subject to serial-dependence effects as seen for many other sensory dimensions. These effects are independent of basic stimulus features such as spatial location, but do depend on whether the previous judgment made was regarding the same (visual variance) or a different (visual direction) dimension, and on the level of confidence expressed in previous judgments. Together, these results suggest that visual variance is processed as a more abstract feature of perception, though is subject to the same processes of efficient coding and perceptual stability found for many other perceptual dimensions.

Experiment 1: Serial dependences in variance judgments

Figure 1 outlines the experimental paradigm utilized across all three experiments. We employed random-dot kinematograms (RDKs), which allow independent manipulation of mean and variance, and asked participants to score the randomness of the motion of RDKs using a visual analogue scale. Several variations of this basic paradigm were used to characterize the effects of previous history on variance judgments.

In Experiment 1, we investigated the existence of serial dependence in variance judgments and its relationship to basic features of stimulus presentation, including eccentricity, spatial location, and mean RDK motion direction. Thus, Experiment 1 employed the basic task (variance estimation without further requirements), with separate blocks in which the RDK was displayed in fovea and periphery.

Methods

Stimuli

The stimulus consisted of a cluster of random moving dots (RDK) displayed for 500 ms at a certain eccentricity (0° or 20°; see later) over a dark-gray background (3.92 cd/m^2). The cluster spanned 5° of visual angle (°va) along the horizontal and vertical dimensions and comprised 100 light-gray dots (diameter = 0.11°va, luminance = 43.14 cd/m^2) moving along a straight trajectory at a rate of 2 pixels/frame (8.45°va/s). The initial position of each dot was uniformly randomized (excluding overlap with other dots), and its coordinates were updated per frame by a trigonometric calculation based on the individual dot's angular-motion direction, re-entering the cluster from the opposite side if it reached a boundary. Each dot's motion direction was extracted from a circular Gaussian (von Mises) distribution that varied for each stimulus presentation: Its mean could take any random integer value from 0° to 359°, and its standard deviation was pseudorandomized among six possible values—namely 5°, 10°, 20°, 30°, 40°, and 60°. This parameter, standard deviation of RDK motion (StD), is the dimension of interest in this experiment.

Procedure

The experimental session comprised a practice block with 72 trials and eight experimental blocks with 60 trials each. The practice had a double purpose: familiarizing participants with the scoring process and the scale used in the experiment, and training maintenance of centered gaze fixation. A broader range of StD values was presented during the practice block compared to the experimental blocks (1°, 10°, 20°, 36°, 60°, and 90°). In both practice and experimental blocks, participants had to score the randomness (variance) of the motion of the RDK using a visual analogue scale (see Figure 1) by adjusting the position of a sliding bar with the mouse. During the practice block, feedback was provided after each response by showing the correct response (score corresponding to the veridical variance) on an additional scale which appeared below the one employed by the participant, after the participant's response. For simplicity, the scale was a linear translation of the StD numeric values ranging from 0° (left end) to 90° (right end). Within this block,

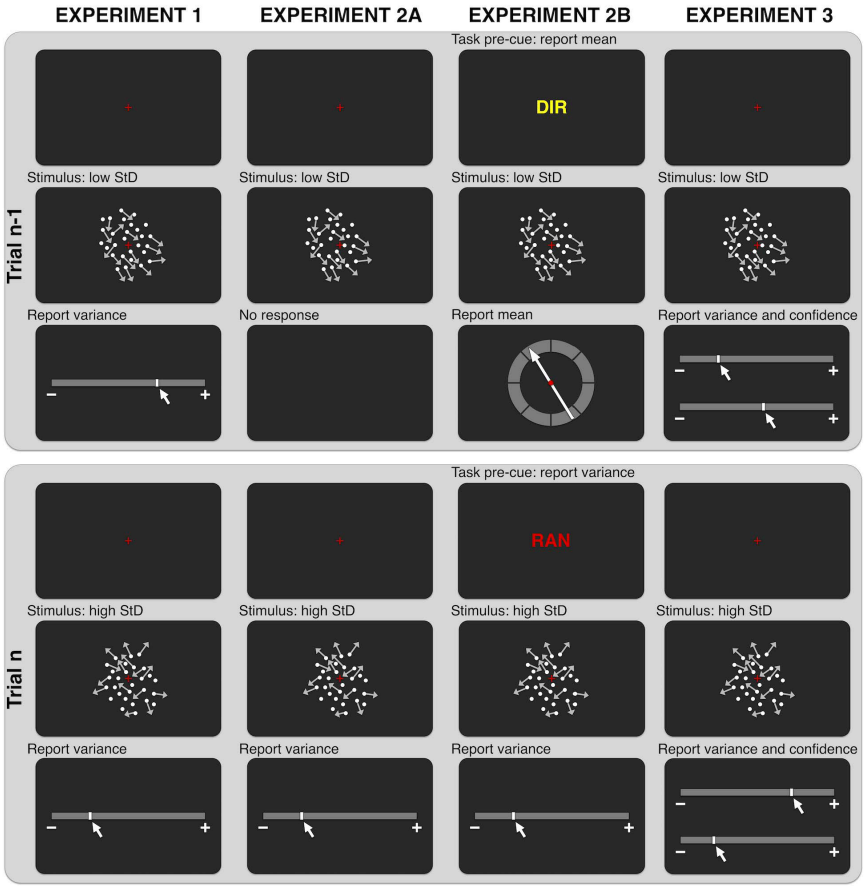


Figure 1. Experiments 1–3: Structure. In all experiments, each trial presented a random-dot kinematogram of a certain mean and variance (standard deviation, StD) in the motion trajectories of its component dots. In the example, Trials $n-1$ and n have low and high StD values, respectively. Experiment 1 required variance (StD) reports for each trial, using a visual analogue scale. Experiment 2 interleaved two thirds of trials in which variance reports were required and one third in which either no response (Experiment 2A) or mean trajectory estimation (Experiment 2B) was required. In Experiment 2B the trial type was precued, so that the label "DIR" or "RAN" at the beginning of each trial indicated whether a mean or variance judgment was required for that trial. Experiment 3 required subjective confidence ratings following variance reports, using a similar visual analogue scale.

the first 36 trials were foveal (the stimulus was presented at 0°va eccentricity) and the remaining 36 were peripheral (20°va).

The eight experimental blocks employed the narrower set of StD values detailed under Stimuli (5°, 10°, 20°, 30°, 40°, and 60°) and did not have feedback. Half of the eight blocks were foveal (stimulus presentation at 0°va eccentricity for all trials) and half peripheral (presentation at 20°va along the horizontal axis), equally frequent in the right and left hemifields. The sequence of foveal and peripheral blocks was pseudorandomized for each participant.

Eye tracking was performed during the entire experimental session. Calibration of the eye-tracking

system was performed at the beginning of each block (practice and experimental) using a standard five-point grid, allowing for a maximal average error of 0.5°va.

At the beginning of each trial a red fixation cross appeared at the center of the screen, spanning 1.1°va (horizontally and vertically). Participants were instructed to maintain their gaze on the fixation cross. The RDK stimulus appeared after 1,000 ms, and both the stimulus and the fixation cross disappeared simultaneously at 1,500 ms from trial onset. Immediately after, the response scale and sliding bar were displayed on the screen. The initial position of the bar was randomized for each trial along the whole length of the scale, to exclude the possibility that participants simply repeated the same (response) action on each trial. If the participant failed to respond within 5 s, the next trial started automatically. The intertrial interval was randomized between 250 and 1,000 ms.

On each trial, participants were allowed to correct their gaze position during the first 700 ms if they noticed that their gaze had deviated from the central fixation cross. However, if a deviation (of more than 5°va) occurred between 700 and 1,000 ms, the trial was aborted and restarted. About a third of participants (9/30) were tested with a slightly different procedure, in which a trial abortion led directly to the start of the next trial (after the intertrial interval). This procedure led to the exclusion of more trials from analysis, since poorly fixated trials were not restarted. Importantly, in both cases trials retained for analysis were those in which fixation was maintained during stimulus presentation (1,000–1,500 ms), and no trial was aborted or restarted after stimulus onset at 1,000 ms.

Participants

Participants were recruited through online advertisement and among members of the laboratory. All were over 18 years old and reported normal or corrected-to-normal vision. Every participant signed an informed consent form before taking part and was either awarded 10 course credits or paid £10 for participation. The study was granted ethical approval by the Research Ethics Committee of the University of Sussex.

Apparatus

Experiments were programmed in MATLAB 2012b (MathWorks, Natick, MA) with Psychtoolbox 3.0.10 and displayed on a LaCie Electron 22BLUE II 22-in. monitor with screen resolution of 1,024 × 768 pixels and refresh rate of 60 Hz. Eye tracking was performed with the EyeLink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at a sampling rate of 1,000 Hz, using a level desktop camera mount. Head position was stabilized at 43 cm from the screen with a chin and forehead rest.

Statistical analysis

Statistical analyses (detailed in the Results section) were performed in MATLAB 2016a, R 3.4.2 (http://www.r-project.org), and JASP (Version 0.8.3.1).

Results

Thirty participants (25 female, five male; mean age = 19.0 years, standard deviation = 1.35) participated in this experiment. Except for two members of the laboratory, they were first-year psychology students who volunteered for course credits.

To ensure the validity of foveal and peripheral conditions, trials without centered gaze fixation during stimulus presentation were removed from the analysis: A trial was deemed valid if the participant maintained fixation within 5°va of the center of the screen for over 80% of the stimulus-presentation period (1,000–1,500 ms from trial onset). Invalid trials were removed, as well as all data regarding trial history that involved at least one of these trials; for instance, if Trial n was valid but Trial $n - 3$ was not, Trial n was not included in analyses regarding serial dependence associated to position $n - 3$ or further backwards. A total of 12,480 trials entered the analysis.

Overview of responses

Figure 2A shows the distribution of responses (R_n) for each StD value and visual eccentricity. Showing that participants were able to perceive the different levels of variance presented in the experiment, reports were positively correlated and monotonically increased with stimulus StD for both foveal and peripheral presentations.

To examine the general pattern of variance judgments, we conducted a repeated-measures analysis of variance (ANOVA) on the influence of two within-subject factors—StD in the current trial (StD_n) and eccentricity—on participant's responses. Both main effects and their interaction were significant (sphericity correction was applied by the Greenhouse–Geisser method). For StD_n, the main effect yielded $F(1.825, 45.621) = 473.80$, $p < 0.001$, $\eta_p^2 = 0.950$, in relation to higher reports for larger stimulus StD. For eccentricity the main effect was $F(1,25) = 33.32$, $p < 0.001$, $\eta_p^2 = 0.571$: Peripheral presentation was associated with lower variance reports, with a mean difference of 6.798 (fovea minus periphery), $t(25) = 8.237$, $p < 0.001$, Cohen's $d = 1.615$. The $StD_n \times$ Eccentricity interaction was also significant, $F(2.715, 67.882) = 20.06$, $p <$

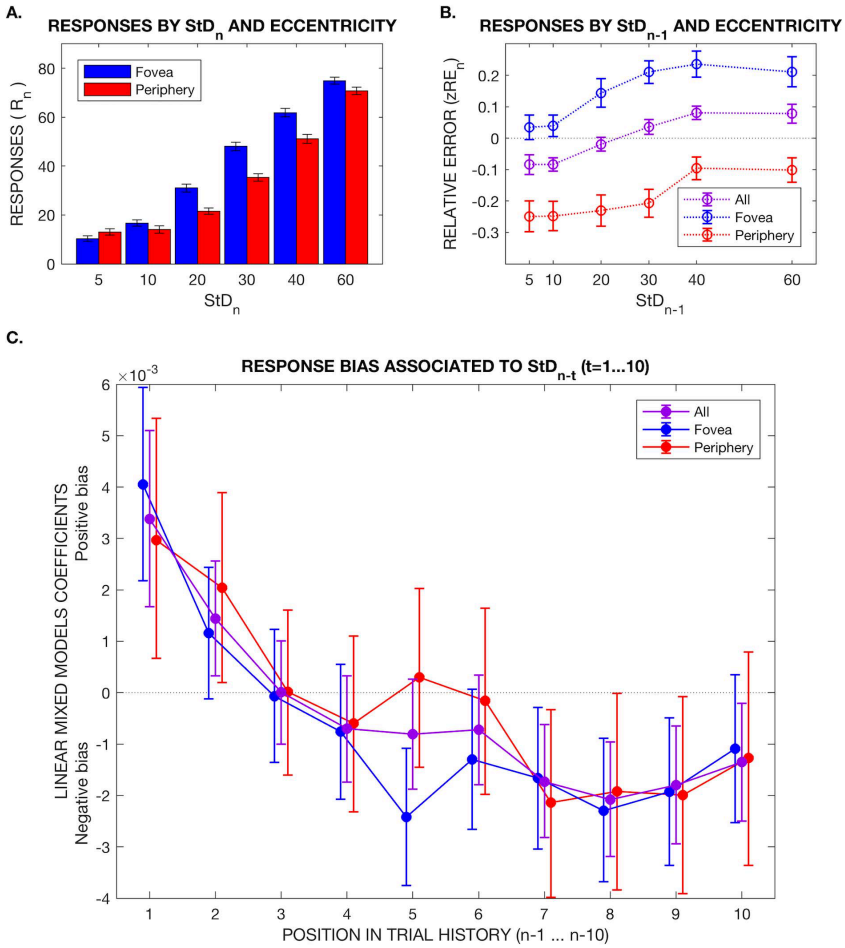


Figure 2. Experiment 1. (A) Distribution of responses by StD_n and eccentricity. The height of the bars represents the mean, and the error bars the between-subjects standard error. (B) Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}). The relative error, defined as $RE_n = (R_n - StD_n)/StD_n$, has been normalized by the distribution of errors provided by each subject for the current StD_n; thus, a positive zRE_n means a larger report in that trial than the participant's average for that stimulus level, and conversely a negative zRE_n indicates a lower-than-average score—that is, sign is not necessarily related to comparison with veridical StD_n, if the participant exhibits a systematic bias for that StD_n. Consequently, plotting zRE_n reports by StD_{n-1} allows examination of any possible bias in relation to previous-trial StD_{n-1}, beyond any unrelated source of bias. The error bars represent the between-subjects standard error. The ascending slope of the plots indicates a positive bias associated with StD_{n-1}, for both foveal and peripheral presentations; relative overestimation occurs for larger StD_{n-1}. (C) Response bias associated with StD presented in recent history. Each data point represents the fixed-effects coefficient estimate (B) in a Bayesian linear mixed-effects model for the association between the StD presented in Trials $n - 1$ to $n - 10$ (StD_{n-t}, $t = 1, \ldots, 10$) and the normalized response error in the current trial. The value of the B coefficient represents the linear slope between the past StD at a certain trial position (StD_{n-t}) and the normalized response error provided in the current trial—that is, the variation (in z scores) observed on the current response (regardless of the presented StD), when StD_{n-t} was increased by 1°. A positive B represents an attractive bias (ascending slope), and a negative B a repulsive bias (descending slope). The error bars depict the 95% credible intervals for the value of the B coefficient.

0.001, $\eta_p^2 = 0.445$, indicating that the difference between foveal and peripheral responses increased for large StD_n values, as shown in Figure 2A. These results were confirmed in a Bayesian repeated-measures ANOVA with the same variables: The full model (both main effects and interaction) was the most explanatory according to the Bayes factor, outperforming the second best (only the two main effects) by a factor of $BF_{full/main} = 1.075 \times 10^6$. These findings (lower responses in periphery than in fovea, especially for large StD_n) seem to relate to a greater regression to the mean exhibited in responses about peripheral stimuli (likely due to worse discrimination between stimulus levels), combined with the fact that the range of the response scale allows for larger errors by overestimation than underestimation.

To further characterize perception of variance throughout the different presented StD levels and confirm the apparently worse performance in the periphery, we examined the dispersion of the responses per StD (σ_R), defined as the standard deviation of the distribution of responses per stimulus level. We conducted a repeated-measures ANOVA on the influence of StD_n level and eccentricity on response dispersion. The main effect for StD_n yielded $F(2.994, 74.840) = 58.426$, $p < 0.001$, $\eta_p^2 = 0.700$, in relation to greater response dispersion for large StD levels, as is common in magnitude-estimation tasks: The value of σ_R was lowest at 9.87 for StD = 5 and steadily increased with StD value until it peaked at 21.03 for StD = 30, remained almost equal ($\sigma_R = 20.51$) for StD = 40, and decreased moderately for StD = 60 ($\sigma_R = 15.98$), probably due to a ceiling effect. As for the main effect of eccentricity on response dispersion, it was $F(1, 25) = 4.165$, $p = 0.052$, $\eta_p^2 = 0.143$, suggesting a trend toward greater response dispersion in peripheral presentations: mean difference = −0.658 (fovea minus periphery), $t(25) = -1.738$, $p = 0.086$, Cohen's $d = -0.339$. Last, the effect of the $StD_n \times$ Eccentricity interaction was $F(3.530, 88.244) = 4.757$, $p = 0.002$, $\eta_p^2 = 0.160$, due to the larger response dispersion in periphery occurring mainly for large StD values. In a Bayesian repeated-measures ANOVA with the same variables as in the frequentist counterpart, the best model was the full model (StD_n, eccentricity, and interaction), which outperformed the second best (with only StD_n) by a factor of $BF_{full/StDn} = 8.747$. In summary, response dispersion increased with stimulus (StD) level and there was a (nearly significant) trend toward greater response dispersion for peripheral presentations, especially at large StDs, suggesting a slightly worse performance at 20°va eccentricity compared to 0°va, in agreement with the previous finding of a greater regression to the mean in peripheral responses.

Variance reports are subject to a positive bias driven by very recent trial history

To characterize the existence of serial dependences in variance reports, we tracked whether the response errors provided by each participant for each StD level were different as a function of the StD level presented in the previous trial (serial dependence in relation to Trial $n-1$) or at positions further backward in trial history (Trial $n-t$). Thus, the response variable in our analyses of serial dependence, unless stated otherwise, is the normalized response error relative to the current stimulus (zRE_n). Response errors, defined as $RE_n = (R_n - StD_n)/StD_n$, are normalized by the distribution of reports provided by each individual for the level of StD presented in the current trial. Thus, zRE_n sums to zero across all trials for a given participant and StD_n level: A negative zRE_n indicates that the participant provided a below-average response in that trial compared to their responses for other physically identical stimuli, while a positive zRE_n indicates an above-average response. Therefore, normalization ensures that the value of the response variable zRE_n is independent of the current StD_n level and of each participant's global scoring biases.

Figure 2B presents the average zRE_n as a function of the previous stimulus (StD_{n-1}), plotted separately by eccentricity. Regardless of generally lower reports at larger eccentricity, a trend toward larger zRE_n for higher StD_{n-1} values is evident for all trials pooled as well as for both foveal and peripheral presentations, as shown by the ascending slope of the three plots (Fovea, Periphery, All). In other words, there was a relative overestimation of the current stimulus when the previous stimulus had a large StD, and a relative underestimation when the previous StD was small, compared to other trials in identical conditions of eccentricity. This indicates a positive (attractive, Bayesian-like) bias driven by Trial $n-1$: Current responses resemble the previous stimulus—serial dependence for visual variance.

To verify this observation, we ran a repeated-measures ANOVA on the effect of StD_{n-1} level (as a within-subject factor) on current variance reports (zRE_n). The effect of StD_{n-1} was statistically significant, $F(3.231, 93.697) = 7.221$, $p < 0.001$, $\eta_p^2 = 0.199$, Greenhouse–Geisser correction applied. The Bayes factor for the inclusion of StD_{n-1} compared to the null model (both of them included participant as grouping variable) was $BF_{inclusion} = 56{,}187.91$, indicating *extreme* (Wagenmakers et al., 2017) evidence for superior explanatory ability of the model that included this term.

Serial dependence in variance does not depend on other stimulus properties (visual eccentricity, spatial location, or ensemble mean): Having established the existence of a positive serial dependence exerted by the previous trial, we sought to ascertain which properties of the

Model	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	Error %
(a) Eccentricity					
Null model	0.200	2.747×10^{-30}	1.099×10^{-29}	1.000	
StD_{n-1}	0.200	1.229×10^{-29}	4.918×10^{-29}	4.476	0.626
Eccentricity	0.200	0.004	0.015	1.385×10^{27}	1.303
StD_{n-1} + Eccentricity	0.200	0.943	65.900	3.432×10^{29}	3.951
StD_{n-1} + Eccentricity + StD_{n-1} × Eccentricity	0.200	0.053	0.226	1.945×10^{28}	2.392
(b) Spatial location (hemifield)					
Null model	0.200	0.232	1.208	1.000	
StD_{n-1}	0.200	0.481	3.702	2.073	0.464
Hemifield	0.200	0.079	0.343	0.340	1.690
StD_{n-1} + Hemifield	0.200	0.181	0.883	0.780	3.545
StD_{n-1} + Hemifield + StD_{n-1} × Hemifield	0.200	0.028	0.114	0.120	1.387
(c) Mean difference ($n-1$, n)					
Null model	0.200	3.111×10^{-6}	1.244×10^{-5}	1.000	
StD_{n-1}	0.200	0.998	2,550.135	320,978.693	0.570
Mean difference	0.200	4.784×10^{-9}	1.914×10^{-8}	0.002	0.726
StD_{n-1} + Mean difference	0.200	0.002	0.006	502.187	0.815
StD_{n-1} + Mean difference + StD_{n-1} × Mean difference	0.200	8.729×10^{-7}	3.491×10^{-6}	0.281	1.091

Table 1. Serial dependence and stimulus properties—model comparison: Experiment 1, serial dependence (associated with Trial $n-1$) and stimulus properties. Each panel presents the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and one property of interest—eccentricity, spatial location (peripheral blocks only: same or opposite hemifield relative to previous stimulus), and difference in the mean trajectories of the random-dot kinematograms presented in consecutive trials. *Notes*: $P(M)$ = prior probability of each model, assumed to be equal for all; $P(M|data)$ = posterior probability of the model (given the data); BF_M = Bayes factor for the model; BF_{10} = Bayes factor for the alternative hypothesis relative to a null model (expressed by each model). All models include subject.

stimulus presentation might modulate such bias. Previous studies on serial dependence have observed that it appears in the fovea as well as the periphery, and its strength is tuned by spatiotemporal proximity (Fischer & Whitney, 2014). Moreover, if the function of (positive) serial dependence is to promote perceptual continuity (Fischer & Whitney, 2014), it seems reasonable to expect that similarity of other attributes of consecutive stimuli would lead to a stronger influence of the studied feature dimension, especially for two attributes as closely related as ensemble mean and variance.

To test the influence of these properties, we conducted a repeated-measures ANOVA on zRE_n (as dependent variable) with two within-subject factors: StD_{n-1} and each of the features of interest separately (eccentricity, retinal location, and similarity of means).

For eccentricity, both main effects were statistically significant—$F_{\text{eccentricity}}(1, 25) = 31.004$, $p < 0.001$, $\eta^2_{p\ \text{eccentricity}} = 0.554$; $F_{StD_{n-1}}(2.662, 66.556) = 7.029$, $p < 0.001$, $\eta^2_{p\ StD_{n-1}} = 0.219$ (Greenhouse–Geisser sphericity correction)—while the interaction was not, $F(3.789, 94.722) = 1.710$, $p = 0.157$, $\eta^2_p = 0.064$. This result, as indicated by the roughly parallel plots for fovea and periphery in Figure 2B, suggests that while eccentricity influences the absolute value of the current StD response, it does not modulate the serial dependence exerted by StD_{n-1}. To formally test this hypothesis, we turned to Bayesian repeated-measures ANOVA. Table 1a summarizes the comparisons between all competing models. The largest Bayes factor corresponds to the model including both main effects but not the interaction ($BF_{10} = 3.432 \times 10^{29}$), which outperforms the model that also includes the StD_{n-1} × Eccentricity interaction by a factor of $BF_{\text{main/full}} = 17.645$—strong evidence (Wagenmakers et al., 2017) against its inclusion and in favor of the conclusion that while there is an overall difference in reports, there is no difference in serial dependence across eccentricity.

Regarding the influence of spatial location, we analyzed only peripheral presentation blocks, classifying trials according to whether the previous stimulus had been presented on the same or the opposite hemifield as the current one—that is, same presentation location versus a separation of 40°va between consecutive presentations. Results for the model comparison given by a Bayesian repeated-measures ANOVA are presented in Table 1b: The best model in terms of evidence includes only StD_{n-1} ($BF_{10} = 2.073$), while the worst model also includes the hemifield and the StD_{n-1} × Hemifield interaction ($BF_{10} = 0.120$). This indicates moderate evidence against the full model (including interaction) compared to the null, and strong evidence against it when compared to the most explanatory model—that is, the one with StD_{n-1} only ($BF_{\text{full/}StD_{n-1}} = 0.058$). These results support the hypothesis of serial

dependence being unaffected by the spatial location of consecutive stimuli. To confirm the absence of tuning by spatial proximity, we further assessed the strength of serial dependence separately for trials with repeated versus opposite hemifield location with respect to the previous stimulus. Results for these analyses are presented in the Supplementary Materials, section 1 (see also Supplementary Figure S1). While the data of Experiment 1 suggest a nonsignificant trend toward a stronger serial-dependence effect for same presentation locations, these results are not confirmed by the data of other experiments; for Experiment 3, the trend goes in the opposite direction.

Last, we examined the influence of mean RDK direction on serial dependence of variance—specifically, whether the magnitude of the serial-dependence effect (in variance) depended on the successive presentations containing a similar mean direction. With this aim, we binned the absolute difference between the mean RDK directions in the previous and current trial into five categories: ≤36°, 37°–72°, 73°–108°, 109°–144°, and 145°–180°. As before, we conducted a Bayesian repeated-measures ANOVA with two within-subject factors ($StD_{n\ 1}$ and mean difference). As shown in Table 1c, the best model included only $StD_{n\ 1}$ ($BF_{10} = 3.210 \times 10^5$), whereas the model including both main effects and their interaction was the second worst (after the one with mean difference only), with $BF_{10} = 0.281$. The Bayes factor for inclusion of the interaction term indicated extreme evidence against it ($BF_{inclusion} = 3.491 \times 10^{-6}$); this was also the case if the comparison was made between the full model and the model lacking only the interaction ($BF_{main/full} = 1789.55$). This lack of association of mean similarity and serial dependence in variance was further confirmed in a different experiment (detailed in Supplementary Materials, section 2) that used a limited range of mean trajectories, allowing for only four between-trials differences (0°, 35°, 55°, and 90°).

In summary, serial dependence in variance reports is not modulated by low-level properties of the stimulus, including visual eccentricity or spatial location, or associated features such as mean, suggesting that visual variance (operationalized as variance of motion direction) is processed as a feature dimension independent from these properties, at least at the level of perceptual decision making that gives rise to serial dependence.

Positive serial dependence in variance extends up to the latest two trials: Investigations of serial dependence have typically focused on the influence of very recent trial history, examining only the effect of the immediately previous and penultimate trials on reports. We examined serial dependence through trial history by modeling the fixed-effects size of serial dependence while allowing for between-subjects variability, building 10 varying-intercept, varying-slope Bayesian linear mixed-effects models (LMMs) with zRE_n as a dependent variable and $StD_{n\ t}$ ($t = 1, \ldots, 10$) as an independent predictor, with random effects grouped by participant. We chose a uniform prior distribution over the real numbers for the fixed-effects coefficient and for the standard deviation of the by-subject varying intercepts and slopes, and an LKJ prior with shape parameter $\eta = 2.0$ for the random-effects correlation matrices. Unless stated otherwise, analogous priors were established for other Bayesian LMMs reported in this article. Fixed-effects coefficient estimates were largely insensitive to prior selection, as can be seen in the example presented in the Supplementary Materials (section 3, Supplementary Figure S2).

We applied these models to foveal and peripheral blocks separately, as well as to the overall data set. Figure 2C presents the fixed-effects coefficients and 95% credible intervals for the association between past StD (up to Trial $n - 10$) and current report for all trials, as well as per eccentricity. The value of the LMM fixed-effects coefficient estimate for the effect of $StD_{n\ t}$ on zRE_n represents the linear slope for the relationship between the StD presented in Trial $n - t$ and the normalized response error provided in the current trial: in other words, the variation (in z scores) in zRE_n when $StD_{n\ t}$ increases by 1°. Therefore, a positive B coefficient represents an attractive bias: A larger StD in a past trial drives a larger response in the present one, regardless of the current stimulus. Conversely, a negative B coefficient represents a repulsive bias.

The fixed-effects B coefficient estimates for the effect of $StD_{n\ 1}$ and $StD_{n\ 2}$ on zRE_n are positive, indicating an attractive bias. For $StD_{n\ 1}$ (all trials pooled), $B = 0.0034$ [0.0017, 0.0051] suggesting that regardless of the value of StD_n, participants' judgments of visual variance increased by a magnitude of 0.0034 (z score) per 1° increase in previous-trial StD ($StD_{n\ 1}$). The effect of $StD_{n\ 2}$ is weaker but still present: $B = 0.0014$ [0.0003, 0.0026]. To make clear the size of these effects, we can consider absolute responses as an outcome variable (adding the current StD_n and the interaction with $StD_{n\ 1}$ to the models). Here, the increase is 0.0586 (0.0272–0.0892) units per unit of $StD_{n\ 1}$, or an attractive effect of 5.9% toward the previous stimulus, whereas for $StD_{n\ 2}$ the effect size is 0.0242 (0.0006–0.0483), or 2.4%.

Thus, variance judgments at one specific trial (n) are attracted by a small but meaningful amount toward the variance presented in the previous trial ($n - 1$) and, to a lesser extent, the trial before that ($n - 2$). Note that since the initial position of the response bar is randomized for each trial, simple motor routines involved in response execution cannot explain this serial dependence.

Variance reports are subject to a negative bias driven by less recent trial history

Looking past the previous two trials, as shown in Figure 2C, a reversal from positive (Trials $n - 1$ and $n -$

2) to negative B coefficient values is observed for less recent presentations, indicative of a negative (i.e., repulsive, anti-Bayesian) bias: Current responses were *less* similar to the StD presented in those trials, in a manner akin to sensory-adaptation aftereffects (Kohn, 2007; Payzan-LeNestour et al., 2016). This effect started at Trial $n - 4$, peaked at Trials $n - 7$ to $n - 9$ (StD_{n-8}: $B = -0.0021$ [−0.0032, −0.0010]), and started to decline afterward. Similar effect sizes and timescales are observed for foveal and peripheral presentations (see Figure 2C).

In the Supplementary Materials (section 4, Supplementary Figure S3) we present a complete exploration of the evolution of the negative effect in relation to more remote positions in trial history. Evidence for the negative effect starts to fade after Trial $n - 9$ but persists to some extent until Trial $n - 20$.

To confirm that the observed serial effects were truly dependent on trial history, we conducted extensive control analyses exploring potential serial dependences in relation to future presentations (StD_{n+1}) and to shuffled data (see Supplementary Materials, section 5, Supplementary Figure S4). These analyses confirm that only in the true trial history is there evidence for the obtained negative and positive aftereffects, supporting the conclusion that these effects are not simply due to statistical artifacts.

Experiment 2: Processing stages involved in serial dependence in variance reports

In the previous experiment we found evidence for serial dependence in judgments about the variance of RDK stimuli. Specifically, we found two opposite types of bias at different timescales: an attractive, Bayesian-like bias related to the StD of the very recent ($n - 1$ and $n - 2$) trials and a repulsive, negative bias which operates on a longer timescale.

At what level of processing do these serial dependences exert their influence? The nonlocal nature and independence from intertrial similarity in RDK direction mean suggest that attractive serial dependence may not be driven by low-level, sensory processes. However, the specific stages of variance processing at which it arises are yet to be determined (Fischer & Whitney, 2014; Fritsche, Mostert, & de Lange, 2017; John-Saaltink et al., 2016). To address this issue, in this experiment we applied several manipulations to the task to disambiguate the contributions of low-level sensory processes, perceptual decision, and responses to serial dependence of variance judgments.

Experiment 2A aimed to isolate the contribution of response to the serial dependence effect by introducing no-response trials, in order to exclude the influence of physically making a response. However, in this experiment no-response trials were not precued, meaning that a potential contribution of decision making and response preparation during stimulus presentation could not be ruled out. For this reason, in Experiment 2B we employed a (precued) task-switching design to disentangle the contribution of perception and decision processes.

Methods

The methods of these experiments were similar to those of Experiment 1, with exceptions as follows.

Stimuli

All stimuli were presented on the center of the screen (where a fixation cross was displayed), and eye tracking was not performed, as visual eccentricity was not under examination.

Procedure

Experiment 2A had a practice block with 72 trials and 10 experimental blocks with 60 trials each; the same was true for Experiment 2B, except that the practice block was longer (90 trials), due to the additional demands of the task-switching design (see later). In Experiment 2B nine participants (out of 16) performed a session twice as long (10 blocks), due to differential availability of different participants. For both experiments, two thirds of the trials in each block required randomness scores as described for Experiment 1. In Experiment 2A, the remaining one third were no-response trials: After stimulus presentation, instead of the response bar only a blank screen appeared, for a randomized interval between 1,000 and 3,000 ms, after which the next trial started. Participants were told in advance that they should expect a certain number of no-response trials, but they did not know the proportion and these trials were not precued in any way. In Experiment 2B, one third of trials required participants to report the mean direction of the motion of the RDK, by adjusting a rotating arrow with the mouse (see Figure 1). The required task was precued at the beginning of the trial: One three-letter string, either "RAN" (randomness report required) or "DIR" (mean direction report required), was displayed for 1 s before the appearance of the fixation cross. The rest of the trial structure was the same as in Experiment 1 (only the response scale differed in RAN and DIR trials).

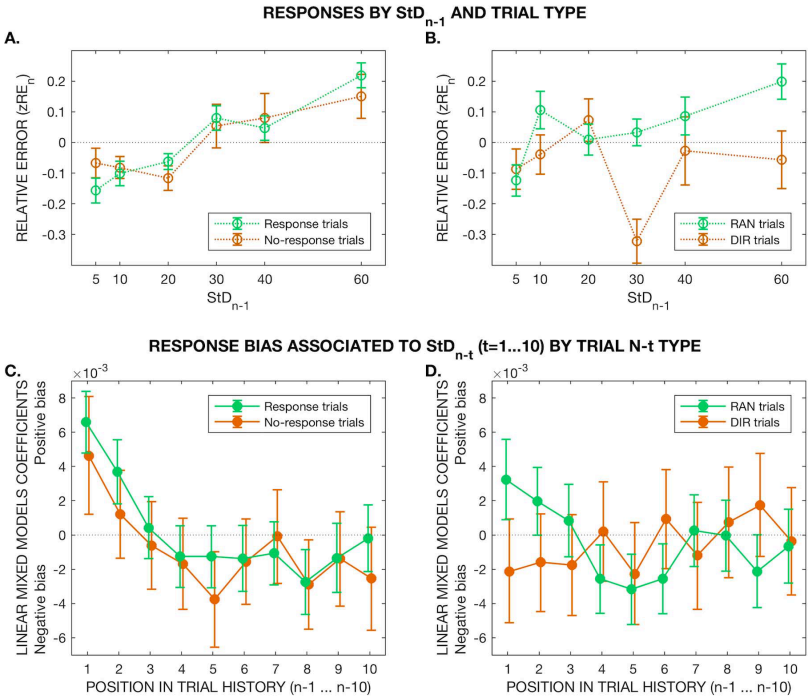


Figure 3. Experiment 2. (A, B) Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by Trial $n-1$ type: (A) response versus no-response (Experiment 2A); (B) RAN versus DIR (Experiment 2B). The error bars represent the between-subjects standard error. Both response and no-response trials are associated with a positive bias by Trial $n-1$, as suggested by the ascending plot lines in (A), whereas in (B), only RAN trials elicit such positive serial dependence. (C, D) Fixed-effects coefficient estimates in 20 Bayesian linear mixed-effects models with StD_{n-t} ($t = 1, \ldots, 10$) as predictor of current response (zRE_n), modeled separately by Trial $n-t$ type: (C) response versus no-response (Experiment 2A); (D) RAN versus DIR (Experiment 2B). Since the dependent variable is the current variance (randomness) judgment, Trial n is always a response (C) or RAN (D) trial. The error bars represent the 95% credible intervals for the true value of the coefficient.

Results

Experiment 2A: Effect of response execution on serial dependence in variance reports

Fifteen psychology students (13 female, two male; mean age = 20.4 years, standard deviation = 5.3) volunteered in exchange for course credits, under the conditions described previously. The total number of trials collected across all participants was 9,000, out of which 3,000 were no-response trials.

Serial dependence of previous StD is not affected by response processes: Figure 3A shows the distribution of normalized variance reports (zRE_n) as a function of the previous trial's StD (StD_{n-1}) and type—that is, whether $n-1$ was a response or a no-response trial. The ascending and roughly parallel plots for each trial ($n-1$) type suggest that serial dependence in relation to StD_{n-1} was similar in magnitude and sign (i.e., attractive effect) regardless of whether Trial $n-1$ was a response or a no-response trial. To formally test this observation, we conducted a Bayesian repeated-measures ANOVA on the effect of StD_{n-1} and Trial $n-1$ type (as within-subject factors) on zRE_n. A comparison of all possible models based on the results of this analysis is shown in Table 2A. The best model includes only StD_{n-1} ($BF_{10} = 2.386 \times 10^6$). There was strong evidence against the inclusion of the $StD_{n-1} \times$ Trial $n-1$ type interaction: $BF_{inclusion} = 0.051$. In a direct comparison between the main-effects model and the full model, the ratio was given by $BF_{main/full} = 10.75$.

Models	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	Error %
(a) Response (Experiment 2A)					
Null model	0.200	3.575×10^{-7}	1.430×10^{-6}	1.000	
StD_{n-1}	0.200	0.853	23.192	2.386×10^{6}	0.393
Trial $n-1$ type (response vs. no-response)	0.200	5.551×10^{-8}	2.220×10^{-7}	0.155	0.772
StD_{n-1} + Trial $n-1$ type	0.200	0.135	0.622	376,281.756	1.178
StD_{n-1} + Trial $n-1$ type + $StD_{n-1} \times$ Trial $n-1$ type	0.200	0.013	0.051	35,151.882	2.835
(b) Dimension-specific judgment (Experiment 2B)					
Null model	0.200	0.012	0.048	1.000	
StD_{n-1}	0.200	0.023	0.095	1.964	0.360
Trial $n-1$ type (variance vs. mean estimation)	0.200	0.109	0.489	9.202	1.348
StD_{n-1} + Trial $n-1$ type	0.200	0.283	1.578	23.921	2.994
StD_{n-1} + Trial $n-1$ type + $StD_{n-1} \times$ Trial $n-1$ type	0.200	0.573	5.371	48.459	2.705

Table 2. Serial dependence and task requirements—model comparison: Experiment 2, serial dependence (associated with Trial $n-1$) and Trial $n-1$ type (task requirements in that trial). The panels present model performance on Experiment 2A and 2B data sets, respectively, according to the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and Trial $n-1$ type (response/no-response in Experiment 2A, RAN/DIR in Experiment 2B). *Notes*: $P(M)$ = prior probability of each model, assumed to be equal for all; $P(M|data)$ = posterior probability of the model (given the data); BF_M = Bayes factor for the model; BF_{10} = Bayes factor for the alternative hypothesis relative to a null model (expressed by each model). All models include subject.

This lack of interaction confirmed that the effect of StD_{n-1} on current reports was independent of response execution.

Figure 3C shows the fixed-effects coefficient estimates and 95% credible intervals for 20 Bayesian LMMs for zRE_n, with StD_{n-t} ($t = 1, \ldots, 10$) as putative predictor, split by Trial $n-t$ type and modeled separately. A similar pattern in terms of effect size and direction can be seen regardless of whether previous trials required response or not: an attractive bias in relation to the latest two trials (weaker for $n-2$), a roughly zero effect of Trial $n-3$, and a reversal toward a negative effect peaking around Trials $n-5$ to $n-9$, with a similar magnitude and timescale as for Experiment 1.

Having established that serial dependence does not arise from response itself, we questioned whether intermediate responses (i.e., responses made in past trials between the current one, n, and a Trial $n-t$ whose serial effect is considered) could affect the degree to which the effect of further trials carried through. For simplicity, we considered only the case of serial dependence related to Trial $n-2$ (for the sake of homogeneity, we limited the analysis to those response trials wherein Trial $n-2$ was also a response trial) and classified the data set according to whether the intermediate trial ($n-1$) was a response or a no-response trial. We ran a Bayesian repeated-measures ANOVA on the effect of StD_{n-2} and Trial $n-1$ type (as within-subject factors) on zRE_n. The best model contained only StD_{n-2} ($BF_{10} = 30.045$), outperforming the full model (two factors and interaction) by a factor of 12.87. However, when the comparison was made between the full model and the equivalent model stripped of the effect of interest (i.e., the $StD_{n-2} \times$ Trial $n-1$ type interaction), the latter outperformed the former by a factor of only $BF_{main/full} = 1.98$. Overall, the Bayes factor for inclusion of the interaction term indicated moderate evidence against it ($BF_{inclusion} = 0.261$), suggesting that the attractive bias related to previous trials is not disrupted (nor boosted) by the participant providing a response on the intermediate trials.

Experiment 2B: Effect of decision on serial dependence in variance reports

Experiment 2A demonstrated that serial dependence in visual variance is not due to response execution; however, as trials were not precued as to whether a response would be required, these results do not disambiguate between perception and decision making (response preparation). Therefore, in Experiment 2B we deployed a precued task-switching design in which participants needed to prepare and respond to two different perceptual tasks: reporting the variance (RAN trials) or the mean (DIR trials) of the motion of the RDK.

Fifteen first-year psychology students (13 female, two male; mean age = 21.4 years, standard deviation = 8.8) participated in this experiment in exchange for course credits, under the conditions already described. In total they performed 7,200 trials, out of which 2,400 were DIR-trials (alternative task).

Serial dependence is related to dimension-specific decision making: We analyzed the data in a similar manner to Experiment 2A, ascertaining the influence of trial type in the observed serial dependence on variance judgments. Figure 3B presents the distribution of variance reports (zRE_n) as a function of StD_{n-1} and Trial $n-1$ type—that is, whether it required a decision

about variance (RAN) or mean (DIR). Only when successive decisions were both regarding variance do we see an ascending slope in relation to increasing StD_{n-1}, suggesting that the attractive bias associated with StD_{n-1} was only exerted if a decision on that dimension had been made.

Table 2b presents a Bayesian repeated-measures ANOVA with StD_{n-1} and Trial $n-1$ type (RAN/DIR) as within-subject factors. The most explanatory was the full model including both main effects and their interaction ($BF_{10} = 48.459$), although the evidence in its favor compared to the model with only the main effects was anecdotal ($BF_{full/main} = 2.026$). However, evidence in favor of the interaction term was larger when taking into consideration all possible models: $BF_{inclusion} = 5.371$, which is moderate evidence. Thus, results point to serial dependence by StD_{n-1} being dependent on which dimension participants had to judge in the previous trial.

We noted that the average time between onsets of consecutive trials was longer if the first was a RAN trial (4.63 vs. 4.48 s, Bayesian pair-samples t test: $BF_{10} = 29.63$). We therefore wondered whether time could be confounding the interaction between StD_{n-1} and Trial $n-1$ response type, since it has been shown to influence serial dependence in previous studies (Bliss, Sun, & D'Esposito, 2017; Fritsche et al., 2017; Kanai & Verstraten, 2005). To test this possibility we defined $time_{n-1,n}$ as the interval between consecutive stimulus onsets, binned into two levels, either below or above the participant's median. This variable was added as a third within-subject factor to the Bayesian repeated-measures ANOVA described in the previous paragraph. We sought to directly compare two explanatory hypotheses for the cause of the observed difference in serial dependence by StD_{n-1} when $n-1$ was a RAN compared to a DIR trial: Trial $n-1$ type or interstimulus time. Thus, we compared the explanatory power of a model with StD_{n-1}, Trial $n-1$ type, and their interaction against a model with StD_{n-1}, $time_{n-1,n}$, and their interaction. The former outperformed the latter by a factor of 105.37, indicating extreme evidence in its favor. Overall, analysis of each separate effect indicated extreme evidence against inclusion of the $StD_{n-1} \times time_{n-1,n}$ interaction ($BF_{inclusion} = 6.668 \times 10^{-4}$). This indicated that the difference between serial dependence driven by RAN compared to DIR trials was better explained by the trial type itself rather than by the intertrial time. There was no support for an independent contribution of time to the observed difference between RAN and DIR trials.

As in previous experiments, we also examined serial dependence within a broader span of trial history. Figure 3D presents the fixed-effects coefficient estimates and 95% credible intervals for the association between StD_{n-t} ($t = 1, \ldots, 10$) and zRE_n, after splitting the data set according to the trial type at each position; thus, the influence of RAN and DIR trials is modeled separately by 20 Bayesian LMMs. As expected from the previous analysis, the positive effect related to StD_{n-1} is present only when those trials required participants to report variance; this is also the case for StD_{n-2}. As for the negative effect appearing at longer timescales, it is clearly present in RAN trials, while for DIR trials, although the credible intervals for the coefficient contain zero at all trial positions (likely due to the smaller number of DIR trials), the negative effect seems to appear as early as Trial $n-1$ ($B = -0.0021$ [–0.0051, 0.0009]), peak at Trial $n-5$ ($B = -0.0023$ [–0.0052, 0.0007]), and decrease afterward. The appearance of a negative serial dependence regardless of the task suggests that it may be sensory in origin—an adaptation aftereffect.

If we ask why there is a serial-dependence effect at $n-1$, we should also ask why there is no such effect at $n-3$. Thus, having established that positive serial dependence arises from feature-specific decision making, we investigated the inverse question: What is the contribution of feature-specific decision making to the fading of positive serial dependence for trials located further away in history? Is this decline affected in a different way by subsequent decisions made on the same, compared to a different, feature dimension? Like for Experiment 2A, we considered all those RAN trials for which Trial $n-2$ was also of type RAN, and examined the association between StD_{n-2} and current report in relation to the intermediate trial's ($n-1$) task. An explanatory role for the $StD_{n-2} \times$ Trial $n-1$ response type interaction would indicate that the intermediate trial type influenced serial dependence related to $n-2$. In a Bayesian two-factor repeated-measures ANOVA, the best model included only Trial $n-1$ response type ($BF_{10} = 41.799$), suggesting that there was no interaction with serial dependence related to StD_{n-2}.

Experiment 3: Influence of confidence in serial dependence

Results of Experiments 1 and 2 indicate that positive serial dependence in visual variance involves mid- to high-level processes, namely decision making about the same feature dimension. In light of this, we questioned how confidence in those decisions modulates serial dependence. We were especially interested in the modulation of the positive (Bayesian-like) bias exerted by very recent trials, in the light of Bayesian accounts of confidence as a measure of the precision of neural representations (Meyniel et al., 2015).

Methods

Stimuli

Stimulus presentation was identical to Experiment 1; we again used foveal and peripheral (20°) presentations, as we considered that the interplay between decision making, confidence, and serial dependence might vary at different degrees of sensory precision.

Procedure

Experiment 3, like Experiment 1, had a 72-trial practice block and eight 60-trial experimental blocks, half of which were foveal and half peripheral. Eye tracking was performed in the same manner as in Experiment 1.

During the response phase of each trial, two identical visual analogue scales were displayed on the screen: the upper one for scoring randomness (variance) and the lower one for confidence (see Figure 1). The initial position of each sliding bar was randomized separately, and the time allowed for responding to both items was 6 s. For data analysis, we obtained the numerical scores as a linear translation from the selected positions; for confidence, the score was expressed as a 0-to-1 proportion of the overall length of the line.

Results

Twenty-two participants (17 female, five male; mean age = 19.6 years, standard deviation = 2.42) volunteered for this experiment. All except for three members of the laboratory were first-year psychology students. As in Experiment 1, trials without valid fixation during stimulus presentation were removed from the analysis, as well as data about trial history of valid trials involving any invalid trial. In total, 8,880 trials were included in the analyses.

Confidence reports correlate with the accuracy and precision of variance judgments

Figure 4A presents the distribution of confidence scores (C_n) plotted by current-stimulus StD (StD_n) and eccentricity. For both foveal and peripheral trials, a trend toward decreasing C_n for larger StD_n is observed, except for the maximal StD (60°). For each StD value, confidence scores are lower in the periphery. To test these observations, we conducted a Bayesian repeated-measures ANOVA on the effect of StD_n and eccentricity (as within-subject factors) on C_n. The best model was the one including both main effects only ($BF_{10} = 6.657 \times 10^{26}$), outperforming the full model with the $StD_n \times$ Eccentricity interaction by a factor of $BF_{main/full} = 9.615$. This indicates that despite the overall lower confidence scores in peripheral blocks, the relationship between different stimulus levels and confidence is the same regardless of eccentricity.

Subsequently we explored whether confidence reports were differentially shaped by response accuracy or precision, and considered the role of eccentricity. Regarding accuracy, we defined error size as the absolute value of the difference between real and reported StD: $E_n = |StD_n - R_n|$. In a Bayesian LMM with C_n as dependent variable and E_n, StD_n, and their interaction as independent variables, C_n reports are inversely associated with error size ($B = -0.0083$, 95% credible interval [−0.0103, −0.0062]) and StD_n ($B = -0.0056$ [−0.0071, −0.0040]) and positively associated with the interaction between both ($B = 0.0003$ [0.0002, 0.0003]). The inverse association between error size and C_n suggests that participants' reports of confidence are, at least in part, grounded in task accuracy. Furthermore, the positive sign of the coefficient estimate for the $E_n \times StD_n$ interaction suggests that confidence tracks relative rather than absolute error: The inverse association between error size (defined as an absolute value) and confidence is weighted down for large StD values. Considering both error size and eccentricity, the negative association with error size remains ($B_{error} = -0.0078$ [−0.0102, −0.0055]), whereas foveal presentations are associated with higher confidence reports independent of task accuracy ($B_{eccentricity} = 0.0510$ [0.0080, 0.0908]). However, the interaction term does not show evidence of a different evaluation of increases in error size in low compared to high eccentricities ($B_{error \times eccentricity} = -0.0013$ [−0.0040, 0.0016]).

As for precision, we calculated the standard deviation of each participant's responses per StD value (σ_R) as a measure of response dispersion. We subsequently modeled confidence by σ_R, StD, and their interaction. As expected, response dispersion shows a negative correlation with confidence: $B = -0.0101$ (95% credible interval [−0.0160, −0.0045]). When we add eccentricity to this model, the main effect for σ_R is close in value ($B = -0.0105$ [−0.0162, −0.0050]), whereas the $\sigma_R \times$ Eccentricity interaction ($B = -0.0003$ [−0.0067, 0.0062]) suggests that the interaction between response dispersion and confidence is similar in fovea and periphery. In summary, our results indicate that confidence is a measure of response precision, and to the extent to which the latter can be considered a proxy for perceptual precision, they are in agreement with Bayesian accounts of metacognition (Meyniel et al., 2015).

Interestingly, we observed a very strong serial dependence for confidence reports. Modeling reported confidence (by a Bayesian LMM) as a function of the report provided in the previous trial (C_{n-1}), we find that the coefficient for the latter is $B = 0.1874$ [0.1445, 0.2307], with $R^2 = 0.3188$. Importantly, if we add the

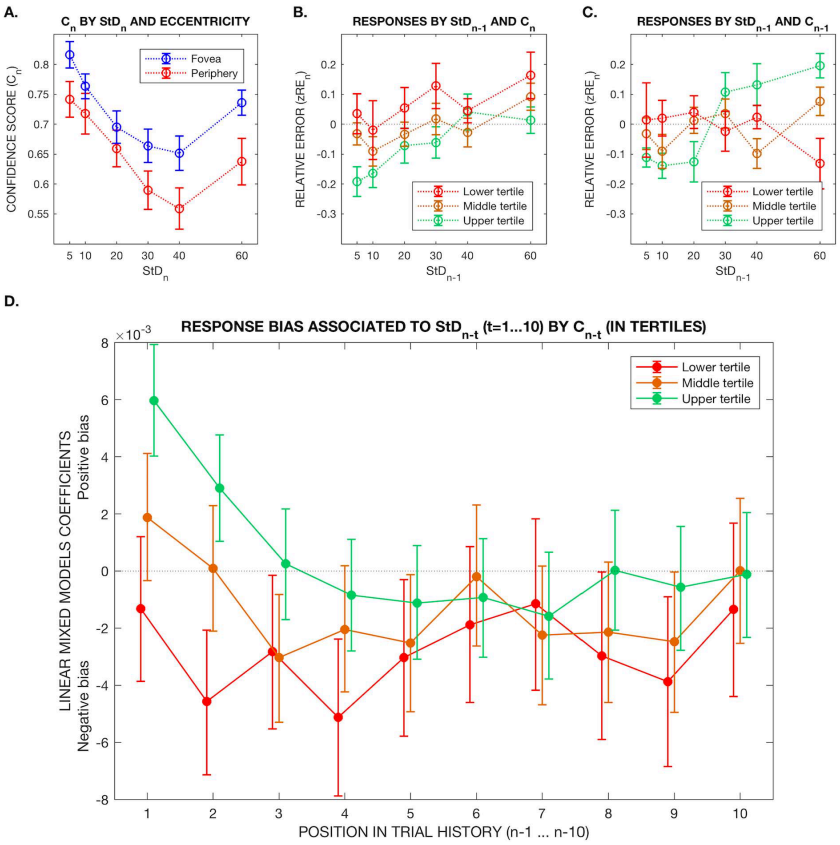


Figure 4. Experiment 3. (A) Confidence scores (C_n) by StD_n plotted separately by eccentricity. (B, C) Normalized relative error in current response (zRE_n) as a function of the StD presented in the previous trial (StD_{n-1}), plotted separately by confidence reported in (B) the current or (C) the previous trial. Confidence scores have been binned into tertiles according to each participant's distribution of reports. The error bars represent the between-subjects standard error. The plots in (B) are all ascending and roughly parallel, indicating that current confidence does not modulate serial dependence by previous-trial StD. Conversely, when considering confidence reported in the previous ($n-1$) trial (C), we observe drastically different slopes: While the high-confidence plot (upper tertile) has a clear ascending slope indicative of a positive bias, the middle-tertile plot is only mildly positive, and the lower-tertile plot is slightly descending, suggesting a negative bias away from low-confidence $n-1$ trials. (D) Fixed-effects coefficient estimates in 30 Bayesian linear mixed-effects models with StD_{n-t} ($t = 1, \ldots, 10$) as predictor of current response (zRE_n), modeled separately by confidence reported in Trial $n-t$ (C_{n-t}), binned into tertiles. The error bars represent the 95% credible intervals for the true value of the coefficient. As suggested for Trial $n-1$ in (C), the size and direction of the bias associated with each trial position depends on the confidence reported in that position, so that the bias will be more negative (or less positive) the lower the reported confidence, within the general trend of an increasingly negative (less positive) bias as we move backward in history.

error size of the previous trial (E_{n-1}) to the model, as well as the $E_{n-1} \times C_{n-1}$ interaction, the coefficient estimate for C_{n-1} has a similar (even larger) value: $B = 0.2197$ [0.1698, 0.2720]. This is also the case when StD_{n-1} is included in the model, suggesting that the serial dependence in confidence scores is due not only to accuracy or attention fluctuating at timescales of several trials, nor to the direct influence of the StD in

Model	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	Error %
(a) Current-trial confidence					
Null model	0.200	0.002	0.010	1.000	
StD_{n-1}	0.200	0.011	0.043	4.437	0.393
C_n	0.200	0.137	0.637	57.150	0.642
$StD_{n-1} + Cn$	0.200	0.841	21.088	349.668	1.213
$StD_{n-1} + Cn + StD_{n-1} \times Cn$	0.200	0.009	0.036	3.738	5.208
(b) Previous-trial ($n - 1$) confidence					
Null model	0.200	0.923	47.768	1.000	
StD_{n-1}	0.200	0.025	0.102	0.027	0.535
C_{n-1}	0.200	0.030	0.123	0.032	1.097
$StD_{n-1} + C_{n-1}$	0.200	8.080×10^{-4}	0.003	8.757×10^{-4}	1.439
$StD_{n-1} + C_{n-1} + StD_{n-1} \times C_{n-1}$	0.200	0.022	0.089	0.024	1.099

Table 3. Serial dependence and reported confidence—model comparison: Experiment 3, serial dependence (associated with Trial $n-1$) and confidence reported in the current and previous trial. The panels report the results of a Bayesian repeated-measures ANOVA on zRE_n, with two within-subject factors: StD_{n-1} and current- or previous-trial confidence. *Notes*: $P(M)$ = prior probability of each model, assumed to be equal for all; $P(M|data)$ = posterior probability of the model (given the data); BF_M = Bayes factor for the model; BF_{10} = Bayes factor for the alternative hypothesis (expressed by each model). All models include subject.

the previous stimulus, but rather may be an expression of response inertia or "confidence leak" as described by Rahnev, Koizumi, McCurdy, D'Esposito, and Lau (2015).

Confidence in a past trial determines the direction of serial dependence in variance reports

According to Bayesian accounts of perceptual decision making, reliance on prior information is greater when the current sensory input is noisy or imprecise, or when the prior itself is highly precise (Cicchini et al., 2014; Petzschner & Glasauer, 2011; Summerfield & de Lange, 2014). Within this framework, confidence is often regarded as a measure of the precision of the sensory signal (Meyniel et al., 2015), a consideration that is in agreement with our data. Thus, we hypothesized that high reported confidence in the current trial (C_n) would decrease any attractive pull toward previous history (with respect to variance judgments), whereas confidence in past trials (C_{n-1}) would have the opposite effect. We further reasoned that such effect of confidence in the past trials would apply mostly to very recent trials, whose information represents a more important contribution when priors are iteratively updated. Indeed, this second hypothesis is in agreement with our observation of a positive bias in variance judgments exerted by only the most recent trials (see Figure 2C for an example).

Figure 4B and 4C depicts zRE_n as a function of StD_{n-1}, plotted separately by confidence in the current (4b) and previous (4c) trial. Confidence scores have been binned into tertiles on a per-participant basis. In Figure 4B, all three plots present an ascending, roughly parallel slope: It appears that serial dependence exerted by Trial $n - 1$ takes place independently of the confidence placed in the current judgment, contrary to our initial hypothesis. However, when we consider the influence of confidence in the previous response, we do see a striking interaction, in line with what would be expected within a Bayesian framework: Low-confidence $n - 1$ judgments do not exert any positive serial dependence—quite the opposite, the plot has a slightly descending slope, pointing toward a negative bias in relation to StD_{n-1}. This slope is mildly ascending for medium confidence and neatly positive only for high-confidence past decisions.

In order to validate these observations, we first performed a Bayesian repeated-measures ANOVA on the effect of StD_{n-1} and C_n (confidence score in the current trial, binned into tertiles) on zRE_n. Results are presented in Table 3a. The best model contains both main effects but not the interaction ($BF_{10} = 349.668$), outperforming the model with the interaction term by a factor of $BF_{main/full} = 93.544$. This provides very strong evidence against the inclusion of the interaction term and indicates that confidence in the current judgment does not modulate serial dependence from the previous trial.

Subsequently we performed an analogous analysis, but with StD_{n-1} and C_{n-1} as within-subject factors. Table 3b presents the results of this analysis. Evidence is in favor of the null model by a large margin (31.25 times more explanatory than the second best, which includes only C_{n-1}). Nevertheless, when we consider the term of interest for our hypothesis, namely the $StD_{n-1} \times C_{n-1}$ interaction, there is strong evidence in favor of its inclusion compared to the model stripped of that effect (including only the two main factors): $BF_{full/main} = 26.989$. Still, because neither competing model was superior to the null model, this result must be taken with caution.

We next asked to what degree confidence in trials located further back in history, up to $n-10$, influenced serial dependence of variance judgments. We split the data set according to the confidence scores reported in each past position (C_{n-t}, discretized into tertiles within each participant's scores), and ran three Bayesian LMMs per position (30 models in total) for the association between StD_{n-t} and zRE_n at each level of past confidence. Figure 4D presents the B coefficient estimates and 95% credible intervals for each Trial $n-t$ ($t = 1, \ldots, 10$). A marked influence of past confidence on the size and direction of serial dependence is observed, such that when high confidence was reported in very recent trials ($n-1$, $n-2$), an attractive pull toward recent StD values is manifest, although this bias fades rapidly, being absent by Trial $n-3$ and thereafter. Note that trials with highest confidence (upper tertile) do not exert a clear, unambiguous negative bias at any point of trial history, although some traces seem to be present from Trial $n-4$ onward. The largest negative bias is driven by low-confidence trials, for which it seems to appear as recently as Trial $n-1$ (although the credible intervals contain zero), becomes unambiguous at $n-2$, and peaks at $n-4$, decreasing afterward—in contrast with the slower buildup of the negative bias seen for past trials with intermediate confidence. Thus, the reversal from positive to negative bias seen in this and previous experiments seems related to the rapid decay of the positive bias of high-confidence trials. As for the negative effect, it seems to appear as early as whenever such competing (positive) bias is not manifest, but fades more slowly. Results were similar when we considered foveal and peripheral blocks separately.

At first glance, the early appearance of the negative effect (after exposure to a single subsecond presentation) and its association with low confidence could suggest that it is at least in part of decisional origin rather than exclusively a product of sensory adaptation. However, some amount of negative bias was observed in relation with past DIR trials in Experiment 2B (trials in which participants were not making a decision on variance). Thus, it seems more likely that the apparent relationship between the negative effect and confidence is due to concealment of the effect in the presence of the positive bias, the latter being associated with high-confidence decision making.

On average, response times for variance reports in low-, medium-, and high-confidence trials were 1.59, 1.46, and 1.30 s, respectively (Bayesian repeated-measures ANOVA: $BF_{10} = 22{,}288$, extreme evidence for the alternative hypothesis), presumably related to subjective trial difficulty. Therefore, we sought to rule out the possibility that the effect of past confidence on serial dependence was related only to the difference in response times, and consequently in interstimulus times. For each trial up to $n-10$, we performed a three-way Bayesian repeated-measures ANOVA for zRE_n (as dependent variable) with three within-subject factors: StD_{n-t}, C_{n-t} (in tertiles), and $time_{n,n-t}$ (time between stimulus onset of Trials $n-t$ and n, binned in two levels with respect to the median). In all cases, the evidence for inclusion of the $StD_{n-t} \times time_{n,n-t}$ interaction was extremely low—that is, the Bayes factor for this specific effect was always below 0.01. This suggests that time was not confounding the reported interaction between confidence and serial dependence.

Time and the additional confidence report might promote an earlier reversal toward negative serial dependence in variance judgments

Experiment 3 had an identical design to Experiment 1 except for the requirement of an additional report (about confidence) per trial. Consequently, an additional difference was introduced: The intertrial time was longer in Experiment 3 that in Experiment 1 (5.06 vs. 3.69 s, Bayesian independent samples t test: $BF_{10} > 6.690 \times 10^7$). As previous work has strongly implicated time between successive stimuli or stimuli and response as critical contributors to serial dependence (Bliss et al., 2017; Fritsche et al., 2017; Kanai & Verstraten, 2005), we sought to take advantage of this circumstance to inquire (post hoc) about the factors that drive the decrease and eventual shift toward negative of the serial-dependence effect as we move backward in trial history.

Figure 5A presents the Bayesian LMM coefficients and 95% credible intervals for the effect of StD_{n-t} ($t = 1, \ldots, 10$) in current variance report as found for Experiments 1 and 3. An extension of this comparison for more distant trial positions is presented and discussed in the Supplementary Materials, section 4 (see Supplementary Figure S3). While the positive bias exerted by StD_{n-1} is similar in magnitude in both experiments ($B = 0.0034$ [0.0017, 0.0051] in Experiment 1, $B = 0.0030$ [0.0018, 0.0042] in Experiment 3), such attraction is still present at StD_{n-2} in Experiment 1 ($B = 0.0014$ [0.0003, 0.0026]) but has virtually disappeared in Experiment 3 ($B = 0.0003$ [−0.0009, 0.0015]). Thus, in Experiment 3 the reversal to negative bias occurs as early as Trial $n-3$ and peaks at $n-5$ ($B = -0.0023$ [−0.0036, −0.0010]), with a similar effect size as the maximum negative bias in Experiment 1, which is seen at $n-8$ ($B = -0.0021$ [−0.0032, −0.0010]). As shown in the Supplementary Materials, negative serial dependences also decline and disappear earlier than in Experiment 1. This earlier buildup of the negative bias could be related to the longer interstimulus intervals in the present experiment: Time might, hypothetically, drive the reversal to repulsive serial effects and posterior fading. Results of Experiment 2B (concerning

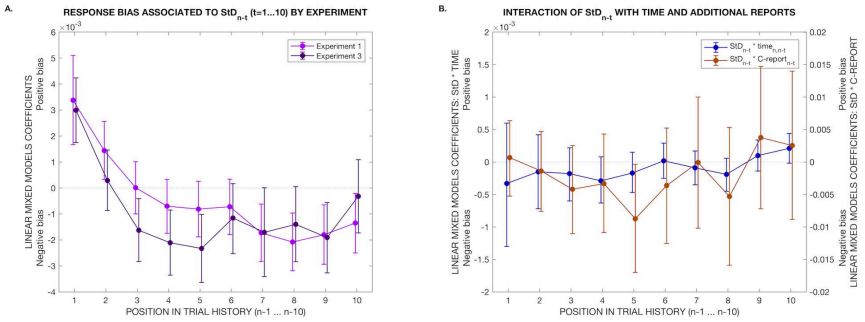


Figure 5. Comparison between Experiments 1 and 3. Both experiments have the same design except for the requirement of a confidence report (in addition to a variance report) per trial in Experiment 3. This also makes the interstimulus time longer, on average, for Experiment 3 compared to Experiment 1. (A) Fixed-effects coefficient estimates in 20 Bayesian linear mixed-effects models with StD_{n-t} ($t = 1, \ldots, 10$) as predictor of current response (zRE_n), with the data of Experiments 1 and 3 modeled separately. The error bars represent the 95% credible intervals for the true value of the coefficient. The shift toward negative coefficient estimates takes place at earlier trial positions in Experiment 3. (B) Fixed-effects coefficient estimates for the $StD_{n-t} \times time_{n,n-t}$ and $StD_{n-t} \times C\text{-report}_{n-t}$ interactions in 10 Bayesian linear mixed-effects models for prediction of zRE_n, with StD_{n-t}, $time_{n,n-t}$, $C\text{-report}_{n-t}$, and all interactions as putative predictors. The variable $time_{n,n-t}$ reflects the time between onsets of the stimuli in Trials $n - t$ ($t = 1, \ldots, 10$) and *n*. $C\text{-report}_{n-t}$ is a binary factor indicating whether confidence reports were made in all trials between $n - t$ and n or in none, regardless of the content of the reports (i.e., the amount of confidence). A negative interaction term with StD_{n-t} indicates a less positive (more negative) serial-dependence effect in relation with longer time or the requirement of an additional confidence report per trial. While credible intervals contain zero in most instances, there is a predominance of negative estimates up to $n - 5$, which could suggest a causal role for both time and the additional confidence report in terms of promoting an earlier reversal of the bias in Experiment 3 compared to Experiment 1.

the effect of DIR trials) and on low-confidence trials in Experiment 3 seem to suggest that the negative bias appears as early as whenever the conditions for the arising of a positive bias are not met. If, hypothetically, positive serial dependence declines with time, the negative effect could become evident in an earlier trial in relation to the longer interstimulus times observed in Experiment 3. Another explanation for the earlier shift toward negative in Experiment 3 would be a disruption of the positive bias caused by the additional confidence report—especially if such Bayesian-like pull is caused by decision processes or depends upon memory to some extent.

To further examine this issue, we pooled all valid trials from Experiments 1 and 3. To ascertain the influence of an additional decision made per trial beside the variance judgment, we defined a binary variable, named $C\text{-report}_{n-t}$, indicating whether or not all intermediate trials between n and $n - t$ had a confidence report in addition to a variance report. Note that the content of the reports (i.e., the amount of confidence) did not affect this definition. When participants missed at least one confidence report in the considered historical span of a certain trial, that trial was excluded from the model, in order to make the comparison unambiguous. Subsequently we built 10 Bayesian LMMs for zRE_n (as dependent variable) in relation with three variables defined at each considered point of trial history, namely StD_{n-t}, $time_{n,n-t}$, and $C\text{-report}_{n-t}$, and all interactions. The fixed-effects B coefficients of the $StD_{n-t} \times time_{n,n-t}$ and $StD_{n-t} \times C\text{-report}_{n-t}$ interactions are plotted in Figure 5B, for Trials $n - 1$ to $n - 10$ as predictors of current variance judgment. A negative interaction coefficient would indicate a comparatively less positive (more negative) serial-dependence effect at that position in relation to longer time or the extra report, respectively.

At all positions, credible intervals for both interaction terms contain zero (except for $StD_{n-t} \times C\text{-report}_{n-t}$ at $n - 5$). However, there is a predominance of negative values for both interaction terms within the recent half of the considered span of trial history, up to Trial $n - 5$. Thus, although results are inconclusive regarding the causes of the different patterns of serial dependence in Experiments 1 and 3, the mostly negative $StD_{n-t} \times time_{n,n-t}$ and $StD_{n-t} \times C\text{-report}_{n-t}$ interactions suggest that both time and the additional confidence report might promote a less positive (more negative) serial dependence in variance and thus contribute to the observed earlier reversal in the direction of the bias. An interesting possibility would be that the dimension-specific, decision-based positive serial dependence is

subject to memory decay as well as a decision-capacity bottleneck. The presented data do not conclusively support a particular interpretation, so future experiments are required to elucidate the relative contribution of time itself and additional judgments in shaping the effects of trial history.

Discussion

The examination of serial dependence provides a valuable window into perceptual processing. In a series of experiments, we applied this approach to visual statistics rather than to individual perceptual features—specifically, to variance, a basic trait in the interpretation of noisy information about complex visual scenes. We found evidence for two opposite serial-dependence effects operating on different timescales: an attractive (positive) bias associated with very recent variance presentations, which is exerted only when a judgment about that dimension was made in the most recent one or two trials and high confidence was placed in that decision, and a repulsive (negative) bias which appears even for the most recent trial history for low-confidence variance presentations but generally becomes manifest several trials into history and persists for at least 10 trials.

Several studies on serial dependence have found a positive (attractive) bias toward recent perceptual history, which is modulated by attention, is enhanced by spatial proximity yet not specific to retinal location, takes place in the fovea as well as the periphery, and fades after 5–15 s but does not require explicit memory (Fischer & Whitney, 2014). While control experiments support the proposition that this effect does not require a motor response, there is an ongoing debate about whether its basis is perceptual or postperceptual: The results of a two-alternative forced-choice discrimination task (with a sample size of three participants; Fischer & Whitney, 2014), a recent behavioral study (Cicchini, Mikellidou, & Burr, 2017), and a V1-based fMRI study (John-Saaltink et al., 2016) have been used in support of a perceptual origin, while another study employing a combination of appearance and performance tasks has made the case for a postperceptual (decisional) source (Fritsche et al., 2017). All four studies examining the mechanistic basis of serial dependence have used a low-level feature like orientation; nevertheless, serial dependence has also been described for high-level features, including facial appearance (Liberman et al., 2014; Xia et al., 2015), relative timing (Roseboom, 2017), and statistical properties such as numerosity (Cicchini et al., 2014) and ensemble mean (Manassi, Liberman, Chaney, & Whitney, 2017).

In our experiments on visual variance (a high-order visual statistic), we found a positive bias that shares many of these characteristics but differs in others: In terms of similarities, it operates on a similar timescale (temporal tuning seems to be slightly shorter for high-level domains, as shown in our data and a study with face perception; Liberman et al., 2014), occurs similarly across presentation eccentricities, and is not related to response execution. It also exhibits other characteristics suggesting that for visual variance, the bias depends on decisional rather than perceptual processes. First, it is entirely independent of retinal location, appearing with similar magnitude for successive stimuli displayed at the same position or at an angular distance of 40°, as shown in the peripheral trials in Experiment 1. Second, it is independent of the closely related statistical property of mean direction (previous studies have highlighted a strong relationship between mean and variance, showing that variance plays an important role in the accuracy and confidence of mean judgments; Fouriezos et al., 2008; Maule & Franklin, 2015). Together, these properties make a low-level, perceptual origin very unlikely. Note that priming of mean judgments by visual variance, as described by Michael et al. (2014), is also independent of the similarity of means and retinal location.

The most compelling argument in favor of a decisional origin for the positive serial dependence in our results is that in a task-switching design, the bias disappears entirely when participants are engaged in a decision about a different feature dimension than variance. This is shown in our Experiment 2B, where participants made decisions about either the variance or the mean direction of the RDK stimuli. This is particularly notable, since mean judgments are strongly dependent on ensemble variance (Fouriezos et al., 2008; Maule & Franklin, 2015), and the stimulus is identical for both tasks. Even so, the possibility remains that the absence of serial dependence in this alternative task may be related to a withdrawal of feature-specific attention (withdrawal of attending to the feature of variance), since serial dependence is enhanced by attention (Fischer & Whitney, 2014). Characterizing the precise effect of attention on serial dependence of variance judgments, and its interaction with task, remains an opportunity for future studies.

From a predictive perception perspective, the fact that only high-confidence trials drive the positive serial dependence may be considered supportive of both perceptual and decisional origin, as a more precise prior would give rise to a stronger reliance on sensory or decisional history (Meyniel et al., 2015). However, an interpretation based on sensory precision might also predict two associations that are not found in our data: an inverse association of positive serial dependence with current-trial confidence, and an inverse associa-

tion with eccentricity, given lower sensory precision in the peripheral field. To the contrary, our experiments strongly support a lack of association of serial dependence with these two factors. In broader terms, serial dependence in variance judgments could be regarded as part of a generic strategy of mirroring or transferring trusted decisions. This explanation could also encompass the negative serial dependence associated with low confidence (as a repulsion away from judgments deemed unreliable); however, the different timescales over which the positive and negative biases operate suggest that they are independent mechanisms rather than two aspects of a confidence-based strategy (Alais, Ho, & Han, 2017).

Finally, several pieces of evidence in our experiments suggest that the positive serial dependence is disrupted by additional decision making, regardless of the domain on which the subsequent decisions operate (variance, mean, confidence). In other words: In all our experiments, the attractive effect of StD_{n-2} stimulus on the current response (which arises only when a high-confidence judgment about variance was made in Trial $n-2$) is much weaker, on average, than that of StD_{n-1}. When inquiring into the factors (interposed between StD_{n-2} and the current response) that might explain this decline, we failed to find any difference based on the *type* of decision that was made in the following trial ($n-1$): In Experiment 2B, the magnitude of the effect of StD_{n-2} (when a variance judgment was made at that point) did not appear to depend upon whether a decision in Trial $n-1$ was made about the variance or the mean of the stimulus. However, if the *number* of interposing decisions was increased, and an additional decision (about confidence) was required in Trial $n-1$, the positive effect of Trial $n-2$ was greatly diminished. This apparent relationship with quantity but not quality of subsequent decisions (made after the one that exerts the bias and before the one that is biased) suggests that serial dependence may be limited by an amodal decision-capacity bottleneck. The apparent fading of the effect with time also points to some sort of memory limitation. Note, however, that these considerations arise from post hoc analyses which revealed only suggestive trends, although the evidence was not conclusive in any case. The factors contributing to the disruption or fading of positive serial dependence in relation with more remote presentations are deserving of further research.

In summary, it is likely that variance-related positive serial dependence is driven by high-level perceptual decision-making processes. In this respect, our findings are in agreement with those of Fritsche et al. (2017), who assert the same for orientation judgments. Those authors propose that working-memory representations are biased toward previous (dimension-specific and task-specific) decisions, a hypothesis that is supported by the potentiation of the bias when several seconds are allowed between stimulus offset and response. A recent study by Bliss et al. (2017) provides converging evidence, reporting that serial dependence is absent at the moment of perception but increases in visual working memory, reaching a maximum when a 6-s delay between stimulus offset and response is imposed (a similar study, however, has reported evidence for serial dependence at the time of perception; Wittmann, Simmons, Aron, & Paulus, 2010). Interestingly, Kanai and Verstraten (2005) have also found a decision-based positive bias on the reported direction of ambiguous motion, appearing only when the stimulus was presented several seconds after the adaptor; they called this effect *perceptual sensitization*.

In our study, participants could respond immediately after the stimulus offset, but due to the relatively long duration of stimulus presentation (500 ms), it is likely that they made an initial decision before that time—as suggested by the results of Experiment 2, wherein the bias exerted by the previous trial was unaffected if a response had not been required (Experiment 2A), but disrupted if a different decision had been indicated by a precue (Experiment 2B). Thus, it is likely that at the moment of response, the representational content produced for the current decision had been already distorted by the memory of previous decisions. More broadly, our results suggest that memory representations of not only the current but also previous perceptual decisions may be subject to similar limitations related to time and informational capacity. While the specific mnemonic processes involved are unclear—our methodology was not designed to operationalize specific instances of memory (such as working memory)—Kiyonaga, Scimeca, Bliss, and Whitney (2017) have noted, in line with our observations, the similarities between serial-dependence effects and well-studied disruptions related to working-memory limitations (such as proactive interference). They suggest that the latter might be a maladaptive aspect of a generally beneficial and widespread brain mechanism for stabilizing internal representations at different levels of processing, including perception, attention, and memory.

Our conclusion of a high-level mechanism of variance processing is also in line with the conclusions of Payzan-LeNestour et al. (2016) regarding variance-driven adaptation aftereffects, which suggest that variance is an abstract property that works independently from its sensory origin and generalizes across domains. Michael et al. (2014) have also proposed variance as an independent property from ensemble average, but suggest that, regarding priming, it operates through feature-specific channels. In our experiments we used a single formalization of variance—dispersion of a dot-motion cloud—so the degree

to which our results will generalize to other variance-related serial dependences requires further investigation.

What are the perceptual or neural mechanisms underlying the observed positive serial dependence? Although this is still uncertain, previous works have proposed exposure-related gain changes or shifts in the neural tuning (Fischer & Whitney, 2014). Furthermore, its behavior resembles that implied by Bayesian frameworks of information processing, in which judgments about a certain dimension are attracted toward prior information. Several studies have recognized that the observed systematic errors in magnitude-estimation tasks, across diverse dimensions, can be well accounted for by assuming an iteratively updated prior, in which recent information is given more weight compared to the overall statistical properties of the environment (Cicchini et al., 2014; Luca & Rhodes, 2016; Petzschner & Glasauer, 2011; Roach, McGraw, Whitaker, & Heron, 2017). Variance-related positive serial dependence indeed shares many characteristics with recursive Bayesian dynamics, including the greater weight of more recent information and the association with high confidence in past trials. Positive serial dependence is probably Bayesian-like in many aspects, but there are some nuances to perceptual decision making that demand further investigation.

The basis of the longer-lasting negative bias is less conclusive, but may be related to adaptation aftereffects, like the variance adaptation described by Payzan-LeNestour et al. (2016). The facts that the negative effect is observed in relation with individual presentations lasting only 500 ms, appears as early as the following trial, and remains even for Trial $n - 9$ could seem unusual for a sensory aftereffect. However, negative aftereffects in response to subsecond stimuli have been described previously (Fritsche et al., 2017; Kanai & Verstraten, 2005), and sometimes lasting for several seconds (Fritsche et al., 2017). Fritsche et al. (2017) have proposed that it is not the stimulus itself but a memory trace that causes the negative aftereffect on orientation. It is likely that the observed relationship between the current trial and a specific trial in history (e.g., $n - 5$) is actually driven by a broader, averaged contextual representation and not by the individual stimuli several trials removed from the present. In our case, as we dealt with a more abstract dimension, we might not consider this high-level aftereffect strictly sensory in the first place (Storrs, 2015). As stated previously, some aspects of this negative bias could point to a decisional component, including its independence of retinal location, predominance in low-confidence trials, and seemingly smaller size when a different decision was required in the past (DIR trials in Experiment 2B; note, however, that the interaction with trial type was not significant). In any case, the line between perceptual and postperceptual aftereffects may be blurred with respect to statistical properties (Payzan-LeNestour et al., 2016; Storrs, 2015).

Some previous studies on different features—both low level (namely motion, Kanai & Verstraten, 2005; and orientation, Fritsche et al., 2017) and high level (such as face attributes; Taubert, Alais, & Burr, 2016)—have reported concomitant positive and negative biases exerted by the same stimulus. Kanai and Verstraten (2005) elicited a negative rapid motion aftereffect of sensory origin by a short, subsecond sine-wave luminance grating presented immediately before. However, when the interstimulus interval was long enough (>3 s), a positive bias was elicited instead, in response to the percept and not the low-level sensory signal (as proven by the use of ambiguous motion adaptors). Fritsche et al. (2017) found opposite effects of recent history on orientation judgments exerted by perception (negative bias) and decision (positive serial dependence), very much in line with our findings. They have proposed that each of these effects has a different biological function, namely increasing sensitivity to changes within the current sensory context and promoting perceptual stability. Taubert et al. (2016) have suggested the same duality in their study of serial dependences in face attributes, although in their case positive and negative biases are exerted concomitantly by different high-level features of the same visual stimulus (faces): Stable traits such as gender would be subject to positive biases in order to smooth away noise, whereas negative aftereffects maximizing sensitivity would predominate in changeable attributes such as facial expresion.

In summary, our study on visual variance reveals two opposite intertrial dependences that operate at different timescales and likely arise at different levels of perceptual decision making: a positive serial dependence in relation to high-confidence, dimension-specific decisions, and a longer lasting negative bias of likely sensory origin. Further investigations are needed to elucidate the precise mechanistic basis of variance-related serial dependence, whether it generalizes to other instances of variance, its relationship to other instances of serial dependence, and the extent to which its properties can be modeled within an iterative Bayesian framework.

Keywords: serial dependence, visual variance, ensemble processing, adaptation aftereffects

Acknowledgments

All authors are grateful to the Dr. Mortimer and Theresa Sackler Foundation, which supports the Sackler Center for Consciousness Science and a PhD

scholarship for MS-P. MS-P is additionally supported by the School of Engineering and Informatics at the University of Sussex, AKS by the Canadian Institute for Advanced Research (Azrieli Programme in Mind, Brain, and Consciousness) and the Wellcome Trust (Engagement Fellowship), and AKS and WR by ERC H2020 project TIMESTORM. MS-P, AKS, and WR designed the research; MS-P performed the research and analyzed the data; and MS-P, AKS, and WR wrote the manuscript.

Commercial relationships: none.
Corresponding author: Marta Suárez-Pinilla.
Email: M.Suarez-Pinilla@sussex.ac.uk.
Address: Sackler Center for Consciousness Science and Department of Informatics, University of Sussex, Brighton, UK.

References

Alais, D., Ho, T., & Han, S. (2017). A matched comparison across three different sensory pairs of cross-modal temporal recalibration from sustained and transient adaptation. *i-Perception*, *8*(4), 1–18, https://doi.org/10.1177/2041669517718697.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567, https://doi.org/10.1177/0956797610363543.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131, https://doi.org/10.1016/j.tics.2011.01.003.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, USA*, *106*(18), 7345–7350, https://doi.org/10.1073/pnas.0808981106.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–8, https://doi.org/10.1167/9.12.13. [PubMed] [Article]

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1):14739, https://doi.org/10.1038/s41598-017-15199-7.

Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15):6, 1–24, https://doi.org/10.1167/15.15.6. [PubMed] [Article]

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: "Cost free" color phenomenality outside focal attention. *Psychological Science*, *25*(7), 1394–1403, https://doi.org/10.1177/0956797614532656.

Campbell, F. W., & Maffei, L. (1971). The tilt after-effect: A fresh look. *Vision Research*, *11*, 833–840.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences, USA*, *111*(21), 7867–7872, https://doi.org/10.1073/pnas.1402785111.

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14):6, 1–9, https://doi.org/10.1167/17.14.6. [PubMed] [Article]

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*(2), 289–301.

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition*, *20*(2), 211–231, https://doi.org/10.1080/13506285.2012.657261.

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*, *66*(6), 937–948, https://doi.org/10.1016/j.neuron.2010.05.018.

de Gardelle, V., & Mammasian, P. (2015). Weighting mean and variability during confidence judgments. *PLoS One*, *10*(3), e0120870, https://doi.org/10.1371/journal.pone.0120870.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, USA*, *108*(32), 13341–13346, https://doi.org/10.1073/pnas.1104517108.

Fairhall, A. L., Lewen, G. D., Blalek, W., & de Ruyter van Steveninck, R. R. (2001, August 23). Efficiency

and ambiguity in an adaptive neural code. *Nature*, *412*(6849), 787–792.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*, 738–743, https://doi.org/10.1038/nn.3689.

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics*, *70*(3), 456–464.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201, https://doi.org/10.1038/nn.2889.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, *27*, 1–6, https://doi.org/10.1016/j.cub.2017.01.006.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192, https://doi.org/10.1146/annurev.psych.58.110405.085632.

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, *15*(4):16, 1–11, https://doi.org/10.1167/15.4.16. [PubMed] [Article]

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble procesing for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734, https://doi.org/10.1037/a0013899.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*(7), 663–676, https://doi.org/10.1068/p7719.

John-Saaltink, E. S., Kok, P., Lau, H., & de Lange, F. P. (2016). Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *The Journal of Neuroscience*, *36*(23), 6186–6192.

Kanai, R., & Verstraten, F. A. J. (2005). Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Research*, *45*(2005), 3109–3116, https://doi.org/10.1016/j.visres.2005.05.014.

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, *21*(7), 493–497.

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*, 3155–3164, https://doi.org/10.1152/jn.00086.2007.

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, *24*(21), 2569–2574, https://doi.org/10.1016/j.cub.2014.09.025.

Luca, M. D., & Rhodes, D. (2016). Optimal perceived timing: Integrating sensory information with dynamically updated expectations. *Scientific Reports*, *6*: 28563, https://doi.org/10.1038/srep28563.

Manassi, M., Liberman, A., Chaney, W., & Whitney, D. (2017). The perceived stability of scenes: Serial dependence in ensemble representations. *Scientific Reports*, *7*(1), 1971, https://doi.org/10.1038/s41598-017-02201-5.

Mather, G., Verstraten, F., & Anstis, S. (1998). *The motion aftereffect: A modern perspective*. Cambridge, MA: MIT Press.

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4):6, 1–18, https://doi.org/10.1167/15.4.6. [PubMed] [Article]

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America A: Optics, Image Science and Vision*, *31*(4), A93–A102, https://doi.org/10.1364/JOSAA.31.000A93.

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*, 78–92, https://doi.org/10.1016/j.neuron.2015.09.039.

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences, USA*, *111*(21), 7873–7878, https://doi.org/10.1073/pnas.1308674111.

Morgan, M. J., Chubb, C., & Solomon, J. A. (2014). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, https://doi.org/10.1167/8.11.9. [PubMed] [Article]

Morgan, M. J., Mareschal, I., Chubb, C., & Solomon, J. A. (2012). Perceived pattern regularity computed as a summary statistic: Implications for camouflage. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *279*, 2754–2760, https://doi.org/10.1098/rspb.2011.2645.

Payzan-LeNestour, E., Balleine, B. W., Berrada, T., & Pearson, J. (2016). Variance after-effects distort risk perception in humans. *Current Biology*, *26*, 1–5, https://doi.org/10.1016/j.cub.2016.04.023.

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *The Journal of Neuroscience*, *31*(47), 17220–17229.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664–1680, https://doi.org/10.1177/0956797615595037.

Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences, USA*, *114*(2), 412–417, https://doi.org/10.1073/pnas.1610706114.

Roseboom, W. (2017). Serial dependence across multisensory relative timing tasks. *PsyArxiv*, https://doi.org/10.17605/OSF.IO/6BKDA.

Roseboom, W., Linares, D., & Nishida, S. (2015). Sensory adaptation for timing perception. *Proceedings of the Royal Society B*, *282*, 20142833, https://doi.org/10.1098/rspb.2014.2833.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4):14, 1–17, https://doi.org/10.1167/12.4.14. [PubMed] [Article]

Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13):13, 1–13, https://doi.org/10.1167/14.13.13. [PubMed] [Article]

Spence, M., Dux, P., & Arnold, D. (2016). Computations underlying confidence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 671–682, https://doi.org/10.1037/xhp0000179.

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology*, *6*:157, 151–154, https://doi.org/10.3389/fpsyg.2015.00157.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision-making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*, 745–756, https://doi.org/10.1038/nrn3838.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409, https://doi.org/10.1016/j.tics.2009.06.003.

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913, https://doi.org/10.1177/0956797614544510.

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*: 32239, https://doi.org/10.1038/srep32239.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76, https://doi.org/10.3758/s13423-017-1323-7.

Wittmann, M., Simmons, A. N., Aron, J. L., & Paulus, M. P. (2010). Accumulation of neural activity in the posterior insula encodes the passage of time. *Neuropsychologia*, *48*(10), 3110–3120, https://doi.org/10.1016/j.neuropsychologia.2010.06.023.

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4):11, 1–13, https://doi.org/10.1167/15.4.11. [PubMed] [Article]

Xia, Y., Liberman, A., Leib, A. Y., & Whitney, D. (2015). Serial dependence in the perception of attractiveness. *Journal of Vision*, *15*(12): 1219, https://doi.org/10.1167/15.12.1219. [Abstract]

Ziemba, C., & Simoncelli, E. (2015). Opposing effects of summary statistics on peripheral discrimination. *Journal of Vision*, *15*(12): 770, https://doi.org/10.1167/15.12.770. [Abstract]

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness & Cognition*, *27*, 246–253, https://doi.org/10.1016/j.concog.2014.05.012.

The illusion of uniformity does not depend on the primary visual cortex: evidence from sensory adaptation

The contents of this paper form a large part of Part III of this thesis.

Article

i-PERCEPTION

i-Perception
2018 Vol. 9(5), 1–13

DOI: 10.1177/2041669518800728
journals.sagepub.com/home/ipe
SAGE

The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence From Sensory Adaptation

Marta Suárez-Pinilla, Anil K. Seth and Warrick Roseboom
Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK;
Department of Informatics, University of Sussex, Brighton, UK

Abstract
Visual experience appears richly detailed despite the poor resolution of the majority of the visual field, thanks to foveal-peripheral integration. The recently described uniformity illusion (UI), wherein peripheral elements of a pattern take on the appearance of foveal elements, may shed light on this integration. We examined the basis of UI by generating adaptation to a pattern of Gabors suitable for producing UI on orientation. After removing the pattern, participants reported the tilt of a single peripheral Gabor. The tilt aftereffect followed the physical adapting orientation rather than the global orientation perceived under UI, even when the illusion had been reported for a long time. Conversely, a control experiment replacing illusory uniformity with a physically uniform Gabor pattern for the same durations did produce an aftereffect to the global orientation. Results indicate that UI is not associated with changes in sensory encoding at V1 but likely depends on higher level processes.

Keywords
perceptual uniformity, uniformity illusion, peripheral vision, tilt aftereffect

Date received: 8 March 2018; revised: 13 July 2018; accepted: 22 August 2018

Introduction

Visual experience appears richly detailed despite the poor sensory precision of the majority (periphery) of the visual field. This topic has received considerable recent attention (Cohen, Dennett, & Kanwisher, 2016; Haun, Tononi, Koch, & Tsuchiya, 2017), with debate about the degree to which visual experience is in fact rich and the potential perceptual processes that may contribute to apparent richness. One recent study demonstrated a compelling example of

Corresponding author:
Marta Suárez-Pinilla, University of Sussex Brighton, BN1 9QJ, UK.
Email: M.Suarez-Pinilla@sussex.ac.uk

how the rich detail within the high-precision central visual field alters peripheral perception the uniformity illusion (UI; Otten, Pinto, Paffen, Seth, & Kanai, 2016). UI describes a phenomenon wherein apparent perceptual uniformity occurs when variable sensory stimulation is presented in peripheral vision, while the central visual field is presented with uniform stimuli. UI occurs for a wide variety of perceptual dimensions,

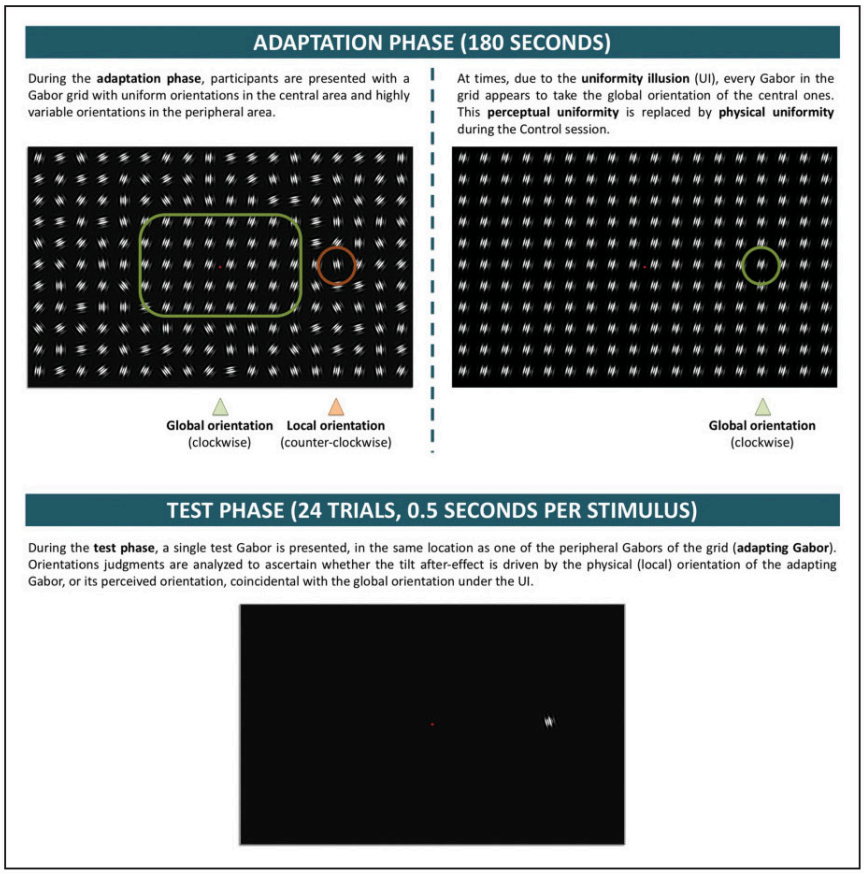


Figure 1. Experimental structure. During the adaptation phase, participants were presented with a Gabor grid wherein the central Gabors had a uniform orientation, while peripheral orientations were heterogeneous. Under UI, perceptual experience was that of a uniform pattern with all Gabors tilted like the central ones. This illusory percept alternated with a nonillusory, nonuniform percept at different times during adaptation. For a specific peripheral Gabor (adapting Gabor), physical and illusory orientation were always in opposition. The control session replicated the phenomenology of the illusion session, replacing perceived with physical uniformity at times in which the participant reported UI in the illusion session. The test phase had 24 trials, wherein participants reported the tilt of a single peripheral Gabor whose location coincided with the adapting Gabor.

including relatively low-level sensory features such as orientation or colour and higher level features such as density (see www.uniformillusion.com for examples).

We sought to examine the mechanisms underlying UI using perceptual adaptation. It is well established that exposure to a specific stimulus magnitude (like an oriented grating) causes perceptual aftereffects (e.g., tilt aftereffect [TAE]; Gibson & Radner, 1937). For visual orientation, perceptual aftereffects have been associated with specific changes in neural coding at the primary visual cortex (V1) and are localized in a retinotopic reference frame (Knapen, Rolfs, Wexler, & Cavanagh, 2010). Here, we utilize the spatial specificity of TAE to examine whether the apparent perceptual uniformity in UI can be attributed to changes in V1-based neural coding for visual orientation. Specifically, we presented participants with Gabor grids wherein the orientation of central elements was uniform, but the orientation of peripheral elements was variable producing UI. At fixed test locations in the periphery of the grid, we presented a physical orientation that differed from the global illusory percept, thus putting local and global orientation in opposition. Following prolonged exposure to global illusory uniformity (UI), we contrasted whether the resultant TAE was consistent with the local, physical orientation or the illusory global orientation.

Methods

Procedure

The experiment had two parts: illusion session and control session. Each session contained six blocks, and each block had an adaptation phase and a test phase (Figure 1). A practice block was run before the illusion session to familiarize participants with UI.

Illusion session. Each block began with an adaptation phase, in which participants were presented with a grid of Gabor patches suitable for producing the UI, affecting the apparent orientation of peripheral elements: All Gabors in the central area had a uniform orientation, whereas orientation of the peripheral Gabors was heterogeneous. Gaze-contingent stimulus presentation ensured that each Gabor was presented to a specific retinal location, as the entire pattern was removed if the participant's gaze deviated from central fixation by more than 1.5 degrees of visual angle (dva) a tolerance threshold equivalent to half the size of each cell of the grid. Adaptation lasted 180 seconds, but because the stimulus was removed when fixation lapsed, actual exposure time could be shorter.

Participants reported the experience of illusory uniformity by pressing a key when all Gabors appeared to take a uniform orientation.

The test phase had 24 trials, separated by a pseudorandom interval of 1,000 to 1,500 ms. In each trial, a single Gabor (test Gabor) was presented for 500 ms at a specific peripheral location, coinciding with the position of a specific Gabor during adaptation (adapting Gabor). Participants reported if the test Gabor was tilted clockwise (CW) or counterclockwise (XCW) from vertical.

Control session. The control session also had six blocks, each built to replicate the phenomenology of a homologous block of the illusion session but replacing illusory for physical uniformity during the adaptation phase.

During the adaptation phase in the illusion session, an empty background was presented whenever the gaze-contingent mechanism removed the adapting pattern. The same pattern of stimulus presentation and removal was replicated in the control session. The stimulus was additionally removed whenever fixation lapsed in the control session. At any other time, the

presentation displayed one of two patterns, differing only in the orientation of peripheral Gabors. The first was identical to the pattern presented in the illusion session and was displayed at times in which the participant had *not* reported UI during adaptation in the illusion session. At times during which the participant had reported UI, the presented pattern was one in which all Gabors had the same *physical* orientation, consistent with the desired illusory orientation during the illusion session. Thus, physical uniformity was inserted at the times in which illusory uniformity had been reported in the illusion session. Participants were not informed that this would occur.

The test phase was identical to that in the illusion session: The location and orientation of the test Gabor in each trial was identical, as well as its test latency (time between the end of the adaptation phase and stimulus onset).

Stimuli

Stimuli were displayed on dark grey background (1.96 cd/m^2). A red fixation dot (8.34 cd/m^2, 0.42 dva diameter) showed constantly on the screen centre.

Gabor patches. Each Gabor consisted of a sine-wave luminance grating with Michelson contrast of 1, 0° phase, and spatial frequency of 1.66 cycles per dva, and a two dimensional Gaussian envelope with a sigma of 0.43 dva.

Adapting pattern. The adapting pattern spanned the entire screen and consisted of a 13×17 grid formed by invisible square cells measuring 3 dva per side (Figure 1). Each Gabor was presented in the centre of each cell. The central area spanned 15 dva horizontally and vertically, encompassing all cells belonging to rows 5 to 9 and columns 7 to 11. All central Gabors had the same orientation, which could be one of two values, each for half the blocks of one session: −15° (global clockwise tilt [GCW]) or 15° (global counterclockwise tilt [GXCW]). The orientations of peripheral Gabors were sampled from a discrete uniform distribution centred on the global orientation and ranging 70° (35° to each side). Thus, mean orientation was the same for central and peripheral Gabors and matched the global orientation perceived under UI.

Two peripheral Gabors of the pattern (adapting Gabors) corresponded to the positions in which the test Gabors would be displayed during the test phase: They were located along the middle (7th) row, at 12.02 dva left and right of the screen centre (columns 5 and 13). Both had the same nonrandomized local orientation, which was the opposite of the global orientation of the block: either 15° (local counterclockwise tilt [LXCW]) or −15° (local clockwise [LCW]).

Henceforth, we give the label adapting condition CX to the presentation pattern wherein the local orientation of the adapting Gabor is clockwise, and the global orientation of the pattern is counterclockwise (LCW, GXCW). Conversely, we will refer to the pattern with LXCW and GCW orientations as adapting condition XC. Both conditions occurred equally frequently during the experiment.

As described above, during the control session, the adapting pattern was replaced by a physically uniform pattern at those times during which participants had reported UI in the illusion session. In these instances, *every* Gabor in the pattern (including the adapting Gabors) took the global orientation.

Test Gabors. A single test Gabor was presented per trial, matching the position of one of the two adapting Gabors. Test Gabors were displayed in the left and right hemifield with equal frequency per block and could take one of eight equally frequent orientations: −12°, −5°,

−2°, −1°, 1°, 2°, 5°, and 12° (negative values indicate CW tilt). Thus, test orientations were always intermediate between global and local orientations (−15°, 15°).

Participants

Participants were recruited through online advertisement, were older than 18 years, and reported normal or corrected-to-normal vision. This study received ethical approval by the Research Ethics Committee of the University of Sussex.

Apparatus

Experiments were programmed in MATLAB 2016a (MathWorks Inc., Natick, MA, USA) and displayed on a LaCie Electron 22BLUE II 22″ with screen resolution of 1,024 × 768 pixels and refresh rate of 100 Hz. Eye-tracking was performed with EyeLink 1000 Plus (SR Research, Mississauga, Ontario, Canada) at sampling rate of 1000 Hz, with level desktop camera mount. Head position was stabilized 43 cm from the screen using chin and forehead rest. Calibration of the eye-tracker was performed at the beginning of each block with a standard five-point grid and a maximal average error of 0.5 dva.

Statistical Analysis

Psychometric curve fitting was performed in MATLAB 2017b, using Palamedes toolbox, version 1.8.1 (Prins & Kingdom, 2009). A cumulative Gaussian curve was fitted by the method of maximum likelihood to the proportion of "XCW" responses per test Gabor orientation, separately for each participant and session/condition (depending on the specific analysis). The threshold (α) for 0.5 proportion of XCW responses and the slope of the curve (β) were free parameters (starting values: $\alpha = 0°$, $\beta = .04$), while guessing (γ) and lapse rate (λ) were fixed at zero.

Bayesian statistics were conducted on JASP (JASP Team, 2017, version 0.8.3.1). For Bayesian t tests, we employed as prior distribution Cauchy(0, $\frac{1}{2}\sqrt{2}$) for two-sided predictions, or a folded Cauchy(0, $\frac{1}{2}\sqrt{2}$) for one-sided predictions (Measure 1 < Measure 2). Likewise, for Bayesian Pearson correlations, we employed a uniform distribution U(−1,1) for two-sided or U(0,1) for one-sided (positive) predictions. For each contrast result, the prior utilized is indicated by the formulated prediction and the subscripts in BF_{10} (two-sided) or BF_{-0} /BF_{+0} (one-sided).

Results

Thirty participants volunteered for the experiment: 23 female; mean age was 21.6 years. To ensure sufficient exposure to the adapting pattern, we excluded blocks wherein the pattern had been displayed for less than two thirds of the adaptation phase (<120 seconds), due to gaze-contingent stimulus removal. In such cases, the corresponding blocks from both control and illusion sessions were removed to maintain balance. This caused exclusion of 32.78% blocks (118/360), including the entire data sets from five participants. Furthermore, because our analyses compared responses across adapting conditions (CX/XC), two additional participants were excluded as all their valid blocks were of only one condition. Results presented here correspond to the remaining 23 participants. Overall, results for all 25 participants with valid blocks were very similar to this counterbalanced sample (see Supplementary Materials, Section S5).

Adaptation Phase

Average exposure time to the adapting pattern per block was 164.13 and 149.47 seconds for the illusion and control sessions: 91.18% and 83.04% of the adaptation phase, respectively. The lower proportion in the control session was expected as pattern removal occurred whenever it had in the illusion session, in addition to times of improper fixation in the control block.

Perceived uniformity was reported, on average, for 43.48 seconds in the illusion session, 26.77% of the time of pattern presentation (minimum 0.55%, maximum 72.23%). The proportion of time of perceived uniformity during the control session was similar to that for the illusion session: 28.41% (minimum 0.59%, maximum 78.42%, Bayesian paired-samples t test: $BF_{01} = 2.733$ anecdotal evidence for the null hypothesis). Physical uniformity in the control session was reported as perceptually uniform 68.13% of the time; by contrast, the nonuniform pattern was reported as uniform only 9.24% of the time. Possibly, presentation of a truly uniform pattern at times shifted a subjective criterion for uniformity by comparison, leading to more conservative reports in the control sessions.

Hypotheses and Measurements

The experiment placed adaptation to illusory and physical orientation in opposition to disambiguate between two competing hypotheses:

(1) The perceived orientation under UI has no effect on tilt adaptation; the TAE is driven solely by the physical orientation of the adapting Gabor.
(2) The global orientation perceived for the entire pattern (including the adapting Gabor) under UI can produce a TAE.

To decide between hypotheses, data were analysed to ascertain the direction of the adaptation-induced bias. We calculated the proportion of XCW reports per test Gabor orientation and obtained the best fitting cumulative Gaussian psychometric curve. The point of subjective equality (PSE) was defined as the test orientation at which 50% reports are XCW. Because CW orientations have (conventionally) negative sign and vice versa, negative PSE indicates a XCW bias, and positive PSE indicates a CW bias.

During the illusion session, a TAE driven by (i.e., away from) the local orientation of the adapting Gabor would imply physical adaptation, while a global-driven TAE would indicate adaptation to illusory orientation. During the control session, both local- and global-driven TAE are compatible with physical adaptation, as the adapting Gabor physically takes the global orientation at times of reported illusory uniformity in the illusion session.

By calculating participants' PSE per adapting condition, we obtained two measurements:

(1) PSE_{CX} and PSE_{XC}. For a local-driven TAE, responses for adapting condition CX should exhibit a XCW bias compared with condition XC ($PSE_{CX} < PSE_{XC}$), and the reverse should happen for a global-driven TAE.
(2) $dPSE = PSE_{CX}\ PSE_{XC}$. We employ this as a summary measure indicating the overall direction of the bias. A negative dPSE indicates a predominance of local-driven TAE ($PSE_{CX} < PSE_{XC}$) consistent with physical adaptation to the local orientation, while a positive dPSE indicates a global-driven TAE, consistent with adaptation to the illusion (or to the physical replication of the illusion during the control session).

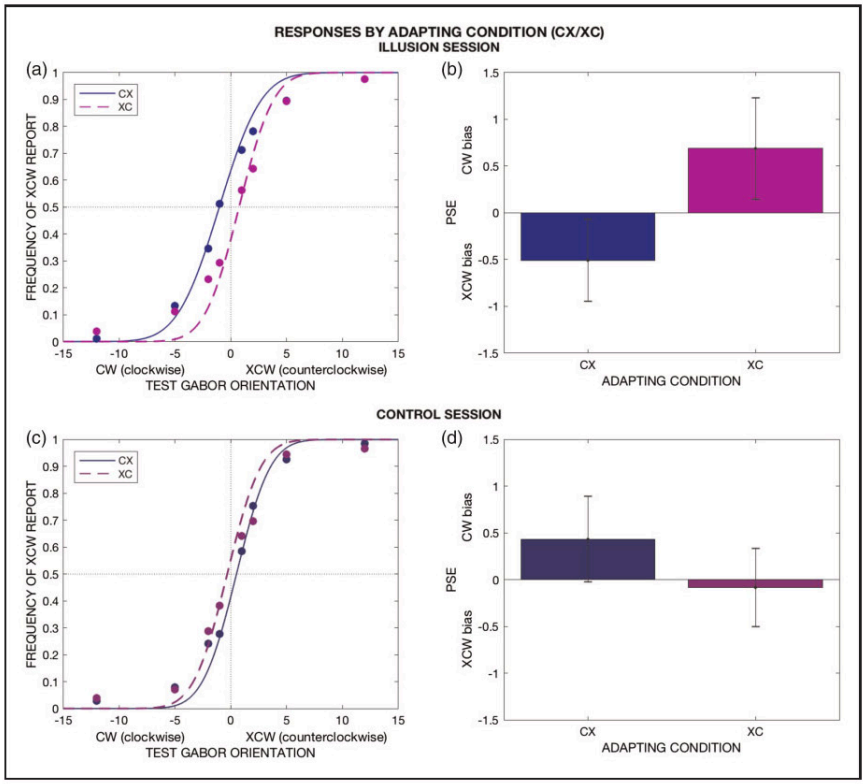


Figure 2. Response patterns by adapting condition. (a) and (c) presents the sample's proportion of counterclockwise (XCW) reports per test Gabor orientation, separated by adapting condition, during the illusion (a) and control (c) sessions. The dotted lines show the best cumulative Gaussian fit for the psychometric curve of each condition, fitted on the sample's pooled data ($N = 23$). (a) and (c) are included for illustrative purposes only, as the PSEs obtained for analysis were computed separately for each participant's data: These results are depicted in (b) and (d), where the bar heights represent the average point of subjective equality (PSE), computed separately per participant, session, and condition. The error bars represent the between-participant standard error. (b) Illusion session. PSEs for both adapting conditions reflect a bias *away* from local orientation (local-driven TAE). (d) Control session. On average (d), responses show a global-driven TAE in CX and are unbiased in XC. These results show that perceived (illusion) and physical (control) uniformity behave differently, suggesting that the TAE is always driven by the physical orientation, even when that orientation is unseen under UI.

TAE Is Driven by Physical, Not Illusory Orientation

Overall effect

Illusion session. Figure 2(a) presents the average proportion of XCW reports per test Gabor orientation during the illusion session, separated by adapting condition (CX or XC). For illustration purposes, it shows cumulative Gaussian curves fitted to the pooled data. However, for analysis, we fitted each participant's responses separately: Individual fits are

detailed in the Supplementary Materials, Section S2. Individual PSEs for each adapting condition are summarized in Figure 2(b). On average, $PSE_{CX} = -0.502°$ and $PSE_{XC} = 0.687°$ reflected a XCW and CW bias, respectively: $dPSE = -1.197°$ ($PSE_{CX} < PSE_{XC}$ Bayesian paired-samples t test: $BF_{-0} = 3.057$) indicated a local, physical-driven adaptation.

Control session. In the control session, the adapting Gabor physically takes the global orientation of the pattern during times of reported uniformity in the illusion session. If adaptation is produced by physical orientation, we should observe a more global-driven TAE compared with the illusion session: $dPSE_{IL} < dPSE_{CO}$. Conversely, if perceived orientation under UI causes adaptation, we should not see a difference between perceived and physical uniformity: $dPSE_{IL} = dPSE_{CO}$.

Results indicate predominance of global-driven TAE during the control session (Figure 2(d)): $PSE_{CX\text{-}CO} = 0.433°$, $PSE_{XC\text{-}CO} = -0.083°$, $dPSE_{CO} = 0.516°$. A Bayesian paired-samples t test comparing dPSE in both sessions was consistent with physical-driven adaptation: $dPSE_{IL} < dPSE_{CO}$, $BF_{-0} = 7.476$. Therefore, the absence of a global-driven TAE in the illusion session was not simply because the global pattern of orientation was insufficient to induce TAE; rather, the illusory (but not the physical) global pattern was insufficient to induce TAE.

The overall predominance of global-driven TAE in the control session, despite presentation of the uniform pattern for only ~27% of time, may be related to a putatively stronger adaptation during this time due to the adjacent Gabors, which then take the global orientation, contributing to the receptive field(s) where the test Gabor will be later presented. Note, however, that the size of each grid cell (3 dva) is larger than the diameter of most receptive fields at V1 (around 1 dva; Bentley & Salinas, 2009), and the relationship between stimulus size and TAE strength is not always intuitive (Harris & Calvert, 1985; Parker, 1972). Another possibility involves extraclassical receptive field effects exerted by the global surround on the adapting Gabor when the latter takes the global orientation (iso-orientation surround suppression; Chen, Chen, Gao, Yang, & Yan, 2015). Whatever the contribution of these effects, they act differently on physical compared with illusory iso-orientation, in the manner expected for low-level processing of the former, but not the latter.

In the Supplementary Materials (Sections S2 and S4), we reanalyse the data set based on raw responses, rather than PSE from fitted curves. Both approaches show the same pattern of results, indicating that choices related to curve fitting and goodness-of-fit of psychometric curves do not significantly affect our analyses.

Time-dependent effect. Overall, responses in the illusion and control session fit the hypothesis that TAE under UI is only driven by physical and *not* illusory orientation. However, in the illusion session, UI is perceived during only ~27% of pattern exposure, on average. Thus, it could be argued that a global, illusion-driven TAE might have been present, but undetected in the overall results, overshadowed by the local-driven TAE at times when UI is not perceived. This possibility seems unlikely because responses in the control session (with uniformity also presented ~27% of time) *do* show an influence of the global-driven TAE. Thus, such a possibility could only hold if the TAE driven by illusory orientation was weaker than that caused by physical orientation. To rule out this possibility, we examined the data from the illusion sessions for evidence of exposure time-dependency of the TAE magnitude. Because the TAE is time-dependent (Patterson, Wissig, & Kohn, 2013), if illusion-driven adaptation was in fact present, we should find evidence for a shift towards more global/less local TAE with longer times of perceived uniformity.

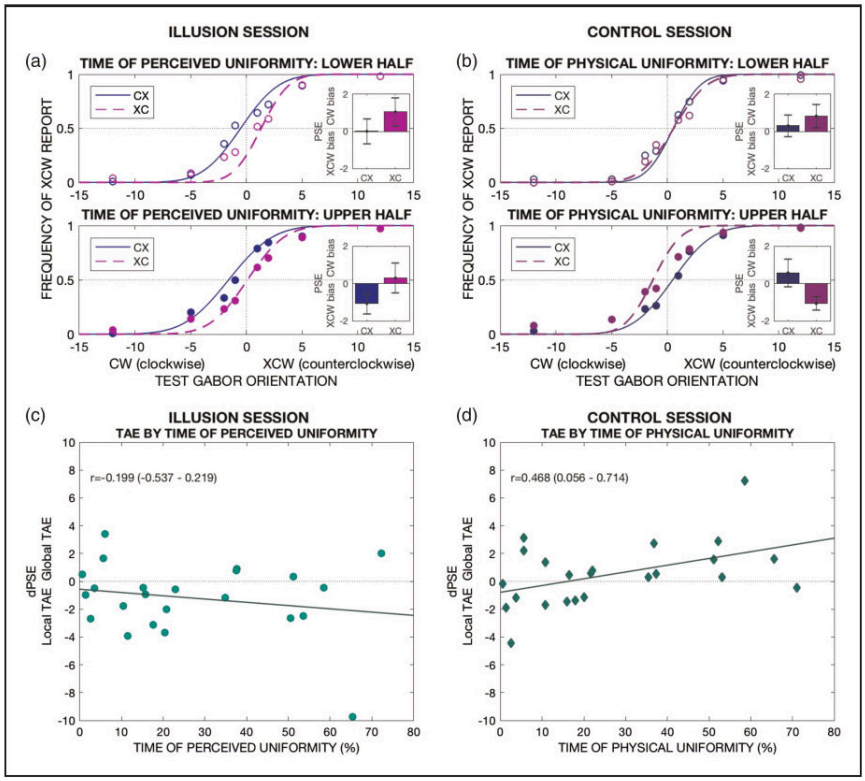


Figure 3. TAE by time of uniformity. Physical, but *not* perceived uniformity, causes a shift towards global-driven TAE in a time-dependent manner. (a and b) Participants are classified into two groups according to whether their average time of uniformity is below (lower half) or above (upper half) the sample's median and depict each group's average responses by adapting condition in the illusion (a) and control (b) session. For illustration purposes, a psychometric function fitted to the pooled data is shown in the main figures; however, all analyses are based on psychometric functions fitted to each participant's data—the group average PSEs of these functions are shown in the insets. In the illusion session, PSEs indicate local-driven TAE regardless of time of perceived uniformity (except for condition CX in the lower half group, which shows no noticeable TAE overall). In the control session, the TAE shifts to global-driven for longer presented physical uniformity. (c and d) The correlation between each participant's average time (%) of uniformity and dPSE during the illusion (c) and control (d) session is presented. For (c), the relationship is established with time of perceived uniformity, and for (d), the relationship is established with time of physical uniformity. Only for the latter (d) do we observe the predicted positive correlation with time, indicative of a shift towards more global TAE. PSE = point of subjective equality; TAE = tilt aftereffect.

Illusion session. If the TAE is driven *only* by physical orientation, in the illusion session, we should expect independence from time of perceived uniformity. Conversely, if the perceived orientation under UI causes adaptation, the response pattern should shift from predominantly local-driven towards more global-driven for longer time of perceived uniformity. We can assess this potential shift by examining dPSE. As stated above,

negative dPSE indicates predominance of local-driven TAE and positive dPSE global-driven TAE. Thus, in the presence of illusion-driven adaptation, dPSE should correlate positively with time of perceived uniformity.

As time measure, we employed the proportion of perceived uniformity (over time of pattern presentation) for conveying the balance between local and (putative) global effects. We analysed the bivariate correlation between time of perceived uniformity and dPSE (Figure 3(c)). Pearson's correlation coefficient and 95% credible intervals were $r = -.199$ ($-.537$, .219), with moderate evidence against a positive correlation: $BF_{+0} = 0.146$.

Therefore, evidence opposed a positive association between time of perceived uniformity and a trend towards more global-driven TAE, thus opposing predictions expected for illusion-based adaptation.

Control session. For the control session, we performed analogous analyses as for the illusion session, but with time of physical instead of perceived uniformity.

Because global uniformity is a physical stimulus in this session, a time-dependent shift from local to global-driven TAE should be expected regardless of the capacity of illusory orientation to induce a TAE. Thus, this control session acts as a sanity check to rule out that the failure to find time-dependency in the illusion session was simply due to insufficient exposure to the global pattern even in the cases of longest time of uniformity.

We performed a Bayesian bivariate correlation between individual average time of physical uniformity and dPSE (Figure 3(d)). Pearson's correlation coefficient was $r = .468$ (95% credible intervals .056, .714), showing moderate evidence for a positive correlation: $BF_{+0} = 5.546$.

Thus, physical uniformity presented for durations equivalent to the reported illusory uniformity was sufficient to observe a shift towards a global-driven TAE.

Discussion

The UI is a striking phenomenon in which experience across the whole visual field is modified by higher precision foveal information, yet its underlying mechanisms remain unknown. Using a version of UI with oriented Gabor patches, we found that UI does not produce an orientation adaptation aftereffect consistent with the illusory percept. Instead, orientation aftereffects only ever followed the (local) physically presented orientation. This suggests that the UI, at least in orientation, arises from higher level (higher than primary visual cortex) perceptual processes.

It has been suggested that the UI may result from predictive processing operations in the visual hierarchy (Otten et al., 2016). In a hierarchical predictive coding scheme, perception arises from the interaction of bottom-up sensory signals with top-down expectations generated in higher cortical areas (Friston, 2005; Rao & Ballard, 1999). Prediction error is determined by the discrepancy between bottom-up sensory signals and the top-down predictions and propagates through the sensory hierarchy to update the internal world model. Although the interplay between neural signatures of sensory adaptation and predictive coding is not fully understood (Symonds et al., 2017), evidence indicates that top-down expectations produce activity changes in the visual cortex also specifically for orientation-selective neurons in V1 (Schummers, Sharma, & Sur, 2005), with adaptation adjusting the relative weight of bottom-up and top-down signals in relation to their precision (Malmierca, Anderson, & Antunes, 2015). Under this framework, UI may be conceptualized as the result of high-precision foveal signals being given more weight in forming perceptual predictions for the presented pattern, possibly in combination with a

prior for perceptual uniformity for the entirety of the visual field. After a period of exposure, adaptation renders low-precision peripheral signals weaker still, until eventually they become unable to overcome the central-based prediction (Otten et al., 2016).

Our results suggest that if UI does result from such predictive operations, the locus of influence of the feedback does not reach primary visual cortex, as illusory uniformity produced no measurable adaptation effect.

What, then, is the neural basis of UI? UI might be an instance of perceptual filling-in, a phenomenon whereby a visual attribute like colour, luminance, or texture is perceived in a region of the visual field even though it only exists in the surround (Komatsu, 2006). However, unlike typical instances of uniform spread of colour or luminance, in our orientation UI, the distinction between background and grid elements persists, and the illusion selectively informs the appearance of the individual Gabors. The process may be similar to texture filling-in or involve texture processing in a broader sense. Notably, several neurophysiological and neuroimaging experiments have reported changes in neural activation in early visual areas that correlate with perceptual filling-in; however, while for colour or luminance, this correlate has been seen at V1 (Hsieh & Tse, 2010), for texture filling-in, it has only been observed at V2 and above (De Weerd, Gattass, Desimone, & Ungerleider, 1995; Komatsu, 2006), in agreement with our results.

UI also exhibits similarities with crowding, as a context-dependent alteration of peripheral perception. Like UI, crowding arises for different low- and high-level dimensions and at several stages of the visual system, involving V2 and above (Whitney & Levi, 2011), for instance, tilt adaptation to the veridical orientation is present for crowded, indistinguishable stimuli (He, Cavanagh, & Intriligator, 1996). Crowding has been likened to texture perception (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). However, as a fundamental difference with crowding, in UI, peripheral phenomenology is not a mixture of adjacent stimuli, but the replacement of peripheral appearance by the traits of sometimes distant foveal elements.

Finally, UI may be due to perceptual inflation, whereby apparent detail in the periphery is not sustained on perceptual content, but due to decisional or metacognitive biases (Odegaard, Chang, Lau, & Cheung, 2018). During the control session in our experiment, where a physically uniform pattern was presented at times, participants were less prone to report UI during presentation of the nonuniform pattern compared with the illusion session: This suggests a shift in decision criterion for uniformity. Importantly, these processes are not exclusive: Possibly both texture processing and perceptual inflation contribute to UI. Further studies may elucidate the precise contribution of the different perceptual mechanisms that underlie foveal-peripheral integration, as demonstrated by UI, and that are central to naturalistic visual experience. However, our results clearly demonstrate that, at least for orientation, these mechanisms do not alter neural coding at the primary visual cortex.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: M. S.-P. is supported by a doctoral scholarship from the Dr Mortimer and Theresa Sackler Foundation and the School of Engineering and Informatics, University of Sussex. W.

R. is supported by the EU FET Proactive grant TIMESTORM: Mind and Time: Investigation of the Temporal Traits of Human-Machine Convergence. A. K. S. is grateful for support to the Dr Mortimer and Theresa Sackler Foundation and to the Canadian Institute for Advanced Research, Azrieli Programme on Mind, Brain & Consciousness.

Supplementary Material

Supplementary material for this article is available online at: http://journals.sagepub.com/doi/suppl/10.1177/2041669518800728.

ORCID iD

Marta Suárez-Pinilla http://orcid.org/0000-0002-6437-3741

References

Bentley, N. M., & Salinas, E. (2009). Neural coding of spatial representations. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (pp. 117–122). Oxford, England: Academic Press.

Chen, C., Chen, X., Gao, M., Yang, Q., & Yan, H. (2015). Contextual influence on the tilt after-effect in foveal and para-foveal vision. *Neuroscience Bulletin*, *31*, 307–316.

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*, 324–335.

De Weerd, P., Gattass, R., Desimone, R., & Ungerleider, L. G. (1995). Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature*, *377*, 731–734.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *360*, 815–836.

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology*, *12*, 453–467.

Harris, J. P., & Calvert, J. E. (1985). The tilt after-effect: Changes with stimulus size and eccentricity. *Spatial Vision*, *1*, 113–129.

Haun, A. M., Tononi, G., Koch, C., & Tsuchiya, N. (2017). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness*, *2017*, 1–4.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337.

Hsieh, P. J., & Tse, P. U. (2010). "Brain-reading" of perceived colors reveals a feature mixing mechanism underlying perceptual filling-in in cortical area V1. *Human Brain Mapping*, *31*, 1395–1407.

JASP Team. (2017). JASP [version 0.8.3.1, Mac OSX Sierra 10.12.6]. Retrieved from https://jasp-stats.org/

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. *Journal of Vision*, *10*, 8.1–8.13.

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Reviews. Neuroscience*, *7*, 220–231.

Malmierca, M. S., Anderson, L. A., & Antunes, F. M. (2015). The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: A potential neuronal correlate for predictive coding. *Frontiers in Systems Neuroscience*, *9*, 19.

Odegaard, B., Chang, M. Y., Lau, H., & Cheung, S. H. (2018). Inflation versus filling-in: Why we feel we see more than we actually do in peripheral vision. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *373*, 20170345.

Otten, M., Pinto, Y., Paffen, C. L. E., Seth, A. K., & Kanai, R. (2016). The uniformity illusion: Central stimuli can determine peripheral perception. *Psychological Science*, *28*, 1–13.

Parker, D. M. (1972). Contrast and size variables and the tilt after-effect. *The Quarterly Journal of Experimental Psychology*, *24*, 1–7.
Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744.
Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *The Journal of Neuroscience*, *33*, 532–543.
Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data*. Retrieved from http://www.palamedestoolbox.org
Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.
Schummers, J., Sharma, J., & Sur, M. (2005). Bottom-up and top-down dynamics in visual cortex. *Progress in Brain Research*, *149*, 65–81.
Symonds, R. M., Lee, W. W., Kohn, A., Schwartz, O., Witkowski, S., & Sussman, E. S. (2017). Distinguishing neural adaptation and predictive coding hypotheses in auditory change detection. *Brain Topography*, *30*, 136–148.
Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168.

How to cite this article

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). The illusion of uniformity does not depend on the primary visual cortex: Evidence from sensory adaptation. *i-Perception*, *9*(5), 1–13. doi: 10.1177/2041669518800728.

CONTRIBUTION TO GRANT REPORT

Here we present a section of a grant report for the EU project TIMESTORM (Action full title: 'Mind and Time: Investigation of the Temporal Traits of Human-Machine Convergence'), forming part of the EU Framework Programme for Research and Innovation HORIZON 2020). Specifically, the presented section corresponds to the Deliverable D4.5: 'Integrated explanation on timely action planning human brain mechanisms', dealing, as the title indicates, with how the human brain performs timing of actions within a complex, sequential and/or multi-agent plan. I was responsible for one of the experiments investigating behavioural aspects of multi-agent collaborative tasks. In particular, this experiment examined the synchronization of physiological measures (heart rate, breathing) between two agents engaged in a joint motor task with a shared goal. However, for logistical reasons, this experiment only reached the preliminary results phase -in particular, the complexity of the design allowed for too many degrees of freedom that could confound our conclusions.

The general experimental aims and structure were designed by Dr. Warrick Roseboom, while the specific details were devised by Dr. Warrick Roseboom, Dr. Darren Rhodes and myself. The programming, data collection, data analysis and writing up of the Results section was done by me. The section of the Deliverable D4.5 that is presented below was largely based on my preliminary results summary, with minor modifications of the text, a new format for the graphs and an introductory paragraph written by Dr. Warrick Roseboom.

Turn-taking task and synchronization of physiological measures

When coordinating with another agent, knowing when to take your turn is paramount to have an efficient interaction. Work in conversational turn-taking shows interpersonal synchrony: respiratory kinematics adjust to the flow and rhythm of a conversation (McFarland, 2001; Warner, Waggener, & Kronauer, 1983). In this study, we use such findings to ask new questions about timely turn-taking. Specifically, we ask whether interoceptive signals such as heart and respiration rate act as markers for interpersonal synchrony in a motor task. We asked pairs of subjects to engage in a task where the aim was to move a cursor to a waypoint (with obstacles). One subject controlled the cursor in the horizontal plane, and the other controlled the vertical plane. As such, to successfully complete the task, the dyads needed to synergistically take turns.

Methods

Six couples of volunteers took part in the experiment, doing as many blocks (of 20 trials each) as they could in 60 minutes. All our results correspond only to blocks with valid physiology recordings. The two participants of each couple switched between mouse and joystick after each block. In all cases the mouse produced horizontal motion and the joystick vertical motion.

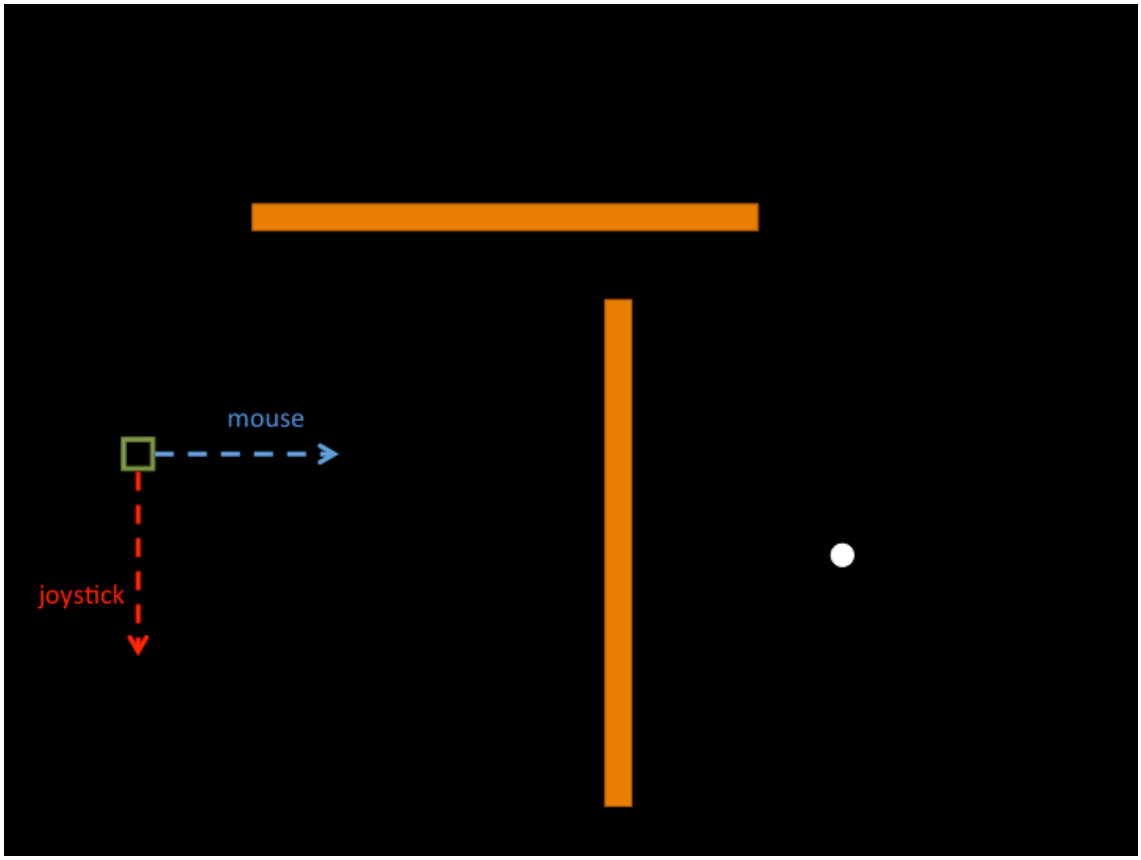


Figure 33. Turn-taking and physiology: Methods. Representation of a screenshot of the collaborative game-like task. The goal of the two-participant team is to take the cursor (represented as a green square) to the white circle, which signals one of the ten waypoints of each trial. Each participant is charged with the control of either a mouse and a joystick, and can only perform movements either in the horizontal (mouse) or the vertical (joystick) direction. After reaching a waypoint, the next one appears on a random location of the screen. To 'win' the trial, participants need to reach ten successive waypoints within a time limit. The trial is 'lost' if the time employed to reach the next waypoint exceeds a certain limit (over 6 seconds) or if the cursor touches the orange barriers (whose position changes each trial).

Overall performance

On average, the rate of successful trials was 60%. Regarding failed trials, the most frequent occurrence (28.95% of fails) was touching one of the obsticles before reaching even the first waypoint. Successful trials are those in which the 10 out of 10 waypoints were reached.

Behavioural measures and performance

Comparing several behavioural measures between successful and failed trials we can report some interesting differences.

Spatial efficiency (the ratio between the distance in straight line from a waypoint to the next, and the actual length of the trajectory followed by the cursor) was *not* greater in successful trials: rather, it was non-significantly lower in those (mixed-effects B coefficient (reference: failed trials): B=-0.064, P=0.069). This was even more marked when considering the joystick (vertical movement) only. However, temporal 'efficiency' (the ratio between the time in which at least one of the devices is producing some movement and the overall time of the trial) was significantly larger in successful trials: B=0.088, P= <.0001. Once more, this difference is larger for the joystick.

When considering movement characteristics, surprisingly, speed (length of the actual trajectory divided by time spent) was much faster in failed trials (B=-134.13, P= 0.0001), once again especially for the vertical movement handled by the joystick. The difference was still more pronounced if we consider only the joystick speed while it is moving without coordination with the other participant (i.e. when the mouse is not moving). Coordination (the fraction of time in which both horizontal and vertical movement is produced simultaneously) was much larger in successful trials: B =0.026, P= <.0001). This was also true (B= 0.036, P=<.0001) when considering the fraction of time in coordinated motion divided by the fraction of time engaged in *any* motion, to avoid confounding by the larger overall temporal 'efficiency' (time engaged in any motion) in successful trials.

Physiology measures and performance

We considered the following physiological measures: heart rate and respiratory rate in each participant of the pair; and the absolute and relative (i.e. divided by the lowest rate

of the two) difference between the heart rates of both participants, as a coarse measure of synchronization. The same analysis was used regarding respiratory rates. Note that the difference is always in absolute values, i.e. we are not concerned with who of the two participants has the highest rate. We also calculated the cross-correlation of the cardiac/respiratory signals of the two participants of the pair, as a measure of synchronization. We consider the maximum (normalized) r coefficient and the lag (shift of a signal along the other) at which this correlation is maximal. We focus especially on the lag: if it is close to zero it implies a good overlap of both signals, and therefore high interpersonal synchronization.

Comparing successful and failed trials, we found that the difference between heart rates of the two participants, as well as respiratory rates, was higher in failed trials. Mixed-effects coefficients for the outcome of the trial in relation to the relative heart and respiratory rate difference are B=-0.071, P = 0.003 (heart) and –B=0.076, P=0.056 (respiration). This result applies to most (but not all) pairs of participants, suggesting a higher synchronization in successful trials.

The lags with maximal cross-correlation between cardiac, and especially respiratory signals are often zero or very close to zero, suggesting a high degree of synchronization, especially concerning breathing. See figure 36, showing the lag values normalized by the cardiac/heart period, i.e., the fraction of the period in which we need to shift the signal of one subject along the other to attain a maximal cross-correlation. Positive lag values indicate subject B's peaks (of blood pressure or respiration) are 'leading': we need to shift B's signal forward to attain maximal overlap with A, and vice versa. However, when assessing the relationship of lag values with performance, we did not find any statistical association. Likewise, we didn't find any association between the maximal r coefficient value and trial performance.

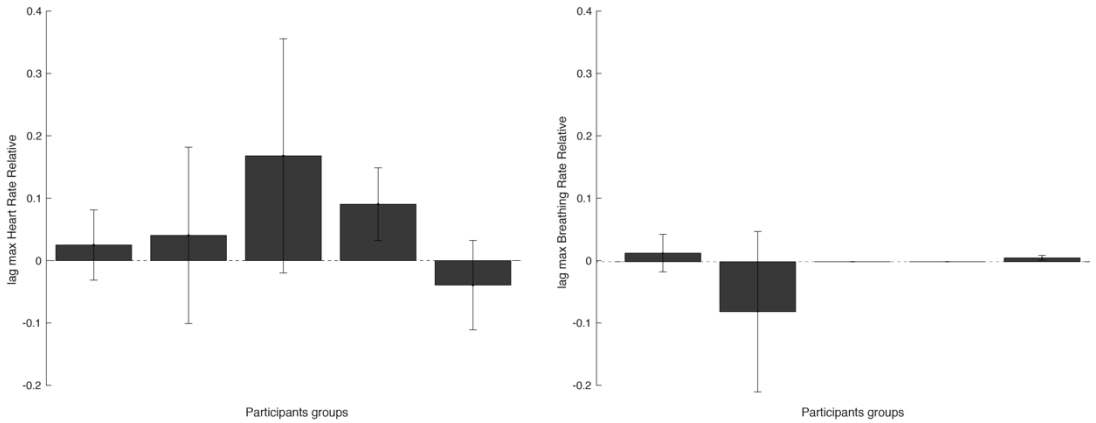


Figure 34. Turn-taking and physiology: Results. Mean lag values normalized by cardiac/breathing period for heart rate (left) and breathing rate (right). Each bar represents a dyad of participants. Values close to 0 suggest a high level of synchronization in the dyad.

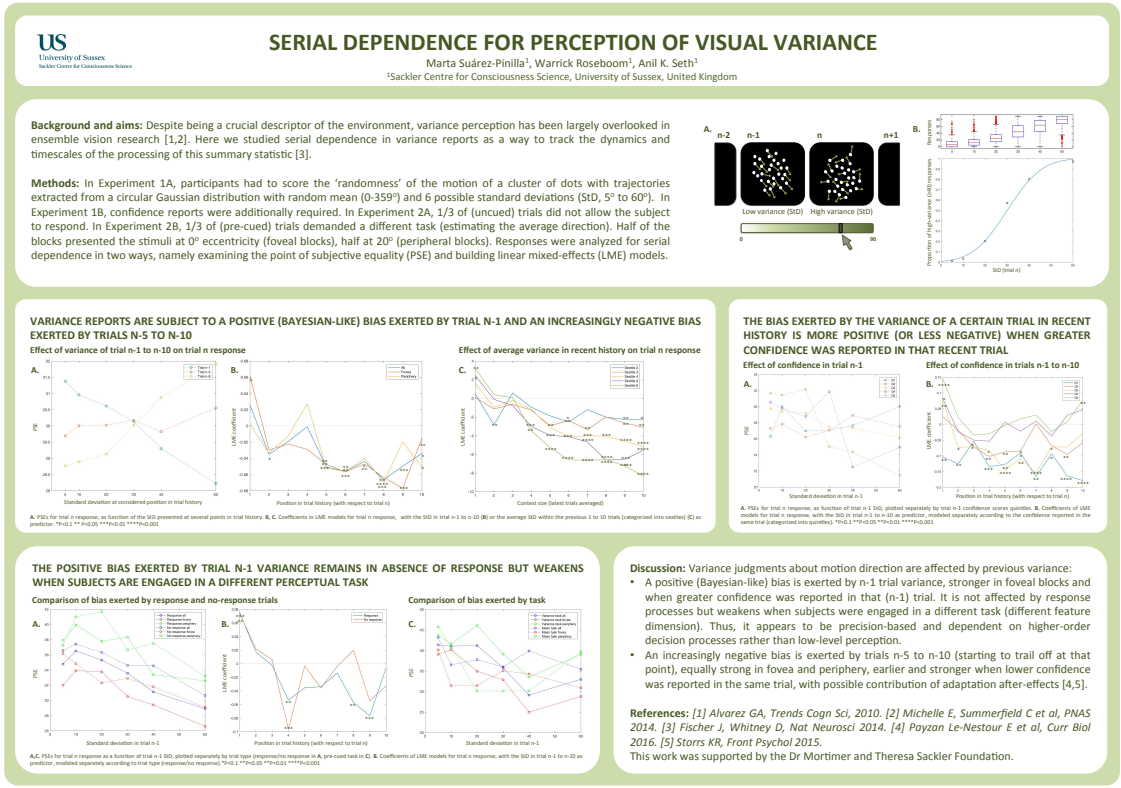
Physiology measures and behavioural measures

In the previous section, we found conflicting indications of a greater cardiac and respiratory synchronization in successful compared to failed trials: the rates where more similar in successful trials, while the cross-correlation was not different. Thus, we set out to ascertain if any of the behavioural differences associated to performance showed in turn a correlation with measures of physiological synchronization, with special attention to coordination of movements. We found that the (relative) difference between the two subjects' respiratory rates was inversely associated with temporal efficiency, speed and proportion of coordinated movement. Thus, more similar respiratory rates were associated with two measures of better coordination (temporal 'efficiency' reflects coordination as it implies less time wasted without any movement, which may reflect indecision about the best strategy or a 'clash of wills' between both subjects). The relationship with speed is less clear. There were no statistical associations with other measures of respiratory synchronization apart from rate difference.

CONFERENCE POSTERS

Serial dependence for perception of visual variance (ECVP 2016)

Presented in the 39th European Conference of Visual Perception (Barcelona, 2016).



University of Sussex
Sackler Centre for Consciousness Science

SERIAL DEPENDENCE FOR PERCEPTION OF VISUAL VARIANCE

Marta Suárez-Pinilla[1], Warrick Roseboom[1], Anil K. Seth[1]
[1]Sackler Centre for Consciousness Science, University of Sussex, United Kingdom

Background and aims: Despite being a crucial descriptor of the environment, variance perception has been largely overlooked in ensemble vision research [1,2]. Here we studied serial dependence in variance reports as a way to track the dynamics and timescales of the processing of this summary statistic [3].

Methods: In Experiment 1A, participants had to score the 'randomness' of the motion of a cluster of dots with trajectories extracted from a circular Gaussian distribution with random mean (0-359°) and 6 possible standard deviations (StD, 5° to 60°). In Experiment 1B, confidence reports were additionally required. In Experiment 2A, 1/3 of (uncued) trials did not allow the subject to respond. In Experiment 2B, 1/3 of (pre-cued) trials demanded a different task (estimating the average direction). Half of the blocks presented the stimuli at 0° eccentricity (foveal blocks), half at 20° (peripheral blocks). Responses were analyzed for serial dependence in two ways, namely examining the point of subjective equality (PSE) and building linear mixed-effects (LME) models.

VARIANCE REPORTS ARE SUBJECT TO A POSITIVE (BAYESIAN-LIKE) BIAS EXERTED BY TRIAL N-1 AND AN INCREASINGLY NEGATIVE BIAS EXERTED BY TRIALS N-5 TO N-10

Effect of variance of trial n-1 to n-10 on trial n response

Effect of average variance in recent history on trial n response

THE BIAS EXERTED BY THE VARIANCE OF A CERTAIN TRIAL IN RECENT HISTORY IS MORE POSITIVE (OR LESS NEGATIVE) WHEN GREATER CONFIDENCE WAS REPORTED IN THAT RECENT TRIAL

Effect of confidence in trial n-1

Effect of confidence in trials n-1 to n-10

THE POSITIVE BIAS EXERTED BY TRIAL N-1 VARIANCE REMAINS IN ABSENCE OF RESPONSE BUT WEAKENS WHEN SUBJECTS ARE ENGAGED IN A DIFFERENT PERCEPTUAL TASK

Comparison of bias exerted by response and no-response trials

Comparison of bias exerted by task

Discussion: Variance judgments about motion direction are affected by previous variance:

- A positive (Bayesian-like) bias is exerted by n-1 trial variance, stronger in foveal blocks and when greater confidence was reported in that (n-1) trial. It is not affected by response processes but weakens when subjects were engaged in a different task (different feature dimension). Thus, it appears to be precision-based and dependent on higher-order decision processes rather than low-level perception.
- An increasingly negative bias is exerted by trials n-5 to n-10 (starting to trail off at that point), equally strong in fovea and periphery, earlier and stronger when lower confidence was reported in the same trial, with possible contribution of adaptation after-effects [4,5].

References: *[1] Alvarez GA, Trends Cogn Sci, 2010. [2] Michelle E, Summerfield C et al, PNAS 2014. [3] Fischer J, Whitney D, Nat Neurosci 2014. [4] Payzan Le-Nestour E et al, Curr Biol 2016. [5] Storrs KR, Front Psychol 2015.*

This work was supported by the Dr Mortimer and Theresa Sackler Foundation.

Sensory mechanisms of perceptual uniformity (ECVP 2017)

Presented in the 40th European Conference of Visual Perception (Berlin, 2017).

University of Sussex
Sackler Centre for Consciousness Science

Sensory mechanisms of perceptual uniformity

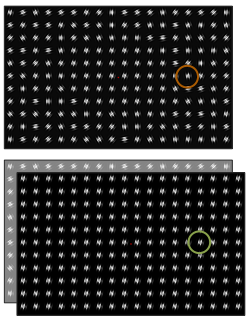
Marta Suárez-Pinilla[1], Anil K. Seth[1], Warrick Roseboom[1]
[1]Sackler Centre for Consciousness Science, University of Sussex, United Kingdom

BACKGROUND: Visual experience appears rich in detail despite the poor performance of the vast majority of the visual field, as a result of the integration of coarse peripheral signals with the information of the comparatively tiny fovea. We examined the mechanisms of this integration by employing the uniformity illusion, in which a pattern with different properties in fovea and periphery uniformly takes the appearance of the fovea[1]. We employed two different perceptual dimensions (orientation and spatial density) to investigate the extent to which the uniformity illusion is associated with changes in sensory encoding.

1. UNIFORMITY ILLUSION ON ORIENTATION

1.1. METHODS

ADAPTATION PHASE (180 SECONDS)



During the **adaptation phase**, participants are presented with a Gabor grid able to produce the uniformity illusion affecting the orientation of the peripheral elements. Eye-tracking ensures that each Gabor patch is received on a specific retinal location, as the pattern is removed if the participant's gaze deviates from the fixation point.

Under the uniformity illusion, all Gabors appear to take the orientation of the central area. Participants report the illusion by holding down a key, allowing us to measure the **time of perceived uniformity.**

In the **control session** (physical uniformity session), the pattern is replaced by a truly uniform Gabor grid at the exact times in which participants reported (illusory) uniformity in the previous session, allowing us to directly compare illusory and physical-driven effects.

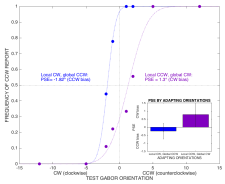
TEST PHASE (0.5 SECONDS PER STIMULUS)

During the **test phase**, a single Gabor patch is briefly presented on a specific peripheral location. Participants report whether it is tilted clockwise or counter-clockwise. We examine which of the two competing orientations causes a tilt after-effect: the **local** (physical) and **global** (illusory).

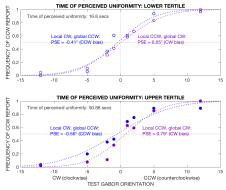
1.2. RESULTS

ILLUSION SESSION: ILLUSORY UNIFORMITY

REPORTS BY ADAPTING ORIENTATIONS

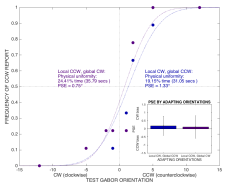


REPORTS BY ADAPTING ORIENTATIONS AND TIME OF PERCEIVED UNIFORMITY

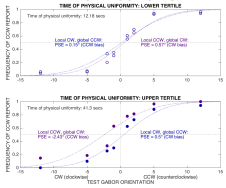


CONTROL SESSION: PHYSICAL UNIFORMITY

REPORTS BY ADAPTING ORIENTATIONS



REPORTS BY ADAPTING ORIENTATIONS AND TIME OF PHYSICAL UNIFORMITY

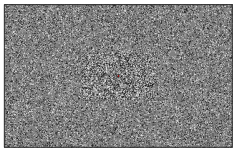


The tilt after-effect follows the local, physically presented orientation rather than the global orientation perceived under the illusion of uniformity. This was not due to insufficient exposure to the global pattern to produce an after-effect as presentation of physical uniformity for the same durations as participant reports of the illusion did produce an after-effect to the global orientation.

2. UNIFORMITY ILLUSION ON SPATIAL DENSITY

2.1. METHODS

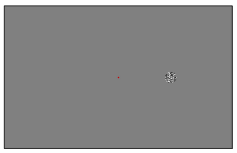
ADAPTATION PHASE (120 SECONDS)

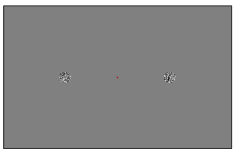


During the **adaptation phase**, participants are presented with a pattern with higher spatial density in the center than the periphery (**Dd**). Some control blocks have uniformly high (**DD**) or low (**dd**) density. In Dd blocks, under the uniformity illusion, phenomenology should be equivalent to pattern DD.

In the **control session**, Dd pattern is replaced by DD at the times in which participants reported the illusion in the previous session.

TEST PHASE (0.5 SECONDS PER STIMULUS)





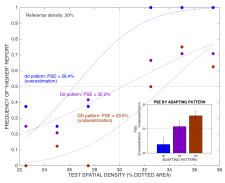
Experiment 2A: Internal reference. Appearance test. Participants report if the density of a peripheral circle is lower or higher than a reference previously learnt. If the uniformity illusion has any effect, responses after exposure to Dd should be more similar to those after DD the longer the illusion has been reported.

Experiment 2B: External reference – 2AFC. Performance (discrimination) test. Participants select, between two peripheral circles, the one with higher density. If the illusion has any sensory effect, responses after Dd should be more similar to those after DD in relation with the time of uniformity.

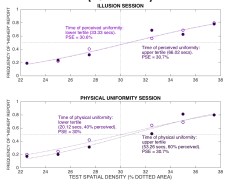
2.2. RESULTS

EXPERIMENT 2A : INTERNAL REFERENCE DENSITY

REPORTS BY ADAPTING DENSITIES (ILLUSION SESSION)

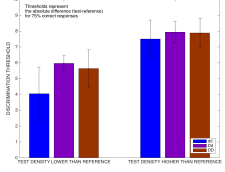


REPORTS BY TIME OF UNIFORMITY (Dd PATTERN)

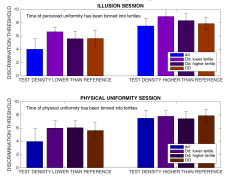


EXPERIMENT 2B : EXTERNAL REFERENCE DENSITY (2AFC TASK)

DISCRIMINATION BY ADAPTING DENSITIES (ILLUSION SESSION)



DISCRIMINATION BY ADAPTING DENSITIES AND TIME OF UNIFORMITY



The intermediate level of responses following exposure to illusory uniformity are consistent with perception having been adapted to the illusory percept. However, as magnitude of effect does not vary with reported time of illusion exposure, this seems unlikely. Rather, the data is consistent with a non-local effect of the high density central display.

DISCUSSION: Experiments performed on two visual domains indicate that the uniformity illusion is not associated with a change in the sensory encoding on a local basis. While it might directly modify more abstract dimensions (such as numerosity, akin to our formalization of spatial density), the time invariance of the effect makes alternative explanations more likely and therefore, suggests that the uniformity illusion arises from high-level perceptual processes. [1] *Otten M, Psychological Science 2016.*

Serial dependence in perception: decision making, metacognition, and awareness (ASSC 2018)

Presented in the 22nd Meeting of the Association for the Scientific Study of Consciousness (Krakow, 2018).

Serial dependence in perception: decision making, metacognition, and awareness

Marta Suárez-Pinilla[1,2], Anil K. Seth[1,2], Warrick Roseboom[1,2]

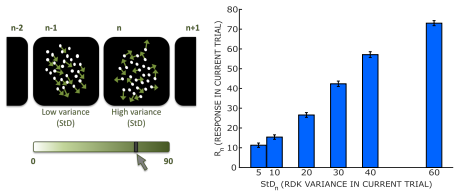
1. Sackler Centre for Consciousness Science 2. Dept. of Informatics, University of Sussex, UK

Introduction

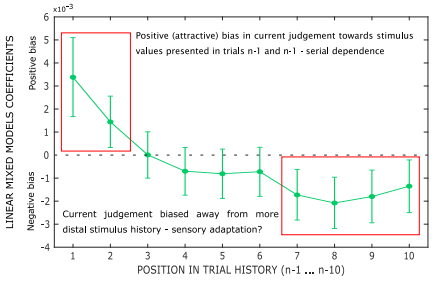
Serial dependence (SD) describes a phenomenon whereby perceptual judgments are attracted towards recent, briefly presented stimuli[1,2]. SD exists for both low and high-level perceptual dimensions, suggesting a central role in shaping visual experience[3].

There is debate about whether SD arises in low-level perceptual and/or higher level decision processing, such as working memory[4]. Furthermore, the relationship of SD to metacognitive processes and perceptual awareness remains unclear.

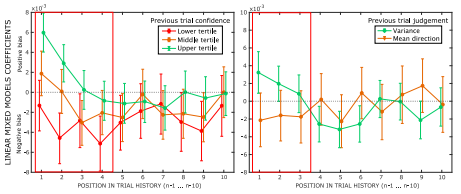
We explored how decision making, confidence, and perceptual awareness constrain SD in visual variance and orientation.



Serial dependence in visual variance



Serial dependence, confidence, and decision making

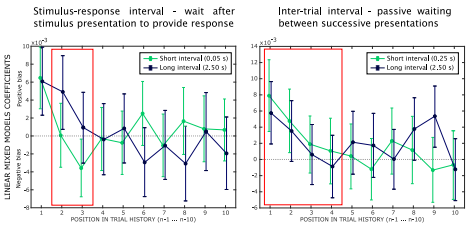


Confidence (left panel):
SD arises only when participants' previous decisions were reported as high confidence. Lowest confidence decisions induce no positive SD bias.

Decision making (right panel):
Positive SD abolished when participants switch between making judgements about the stimulus variance and mean stimulus direction. This is true regardless of whether the required judgement was pre- (shown) or post-cued - ruling out 'attention' related explanations.

Contact: ms822@sussex.ac.uk www.sussex.ac.uk/sackler

Serial dependence of decisions over time



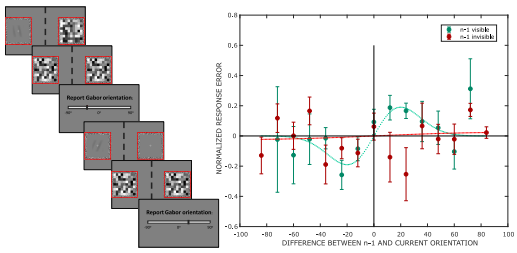
Active maintenance of decision enhances SD (left panel):

When waiting to provide response following stimulus presentation (on the current trial), there is an increased bias towards less recent (n-2 and n-3) presentations[4].

SD decays with time (right panel):
Passively waiting between successive trial presentations (not required to maintain response, etc) reduces the bias towards previous trials.

Serial dependence requires visual awareness

We explored SD in the orientation of a sequence of Gabor patches using a continuous flash suppression paradigm where 50% of presentations were masked.



Characteristic derivative of Gaussian tuning of SD in orientation[4] only arises when n-1 trial was visible. Counter to previous claims[1] and unlike the sensory adaptation seen in illusions like the tilt after effect, SD in orientation is not the result of low level visual processing.

Conclusions

Generation of SD: past stimuli exert SD on subsequent presentations in presence of:
a. High decisional confidence for past decision.
b. Contiguity of feature dimension about which decision is made.
c. Perceptual awareness of past stimulus.

Maintenance of SD: duration of influence of a past stimuli is affected by:
a. Time-related (passive) decline.
b. Active maintenance of current representation in working memory. Longer maintenance of a decision before report increases bias toward past presentations[4].

Serial dependence is consistent with high-level process of iterative (Bayesian-like) biases in decision rather than perception and only occurs in the presence of high decisional confidence and perceptual awareness.

References

1. Fischer J, Whitney D. (2014) .Serial dependence in visual perception. Nature Neuroscience, 17:738-43.
2. Suárez-Pinilla, M., Seth, A.K. & Roseboom, W. (in press). Serial dependence in the perception of visual variance. Journal of vision.
3. Kiyonaga, A., Scimeca, J.M., Bliss, D.P., Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory.Trends in Cognitive Sciences. 21(7):493-7.
4. Bliss, D.P., Sun, J.J., D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. Scientific Reports. 7, 14739.

Acknowledgements

This work was supported by a donation from the Dr Mortimer and Theresa Sackler Foundation and the H2020 FET grant TIMESTORM (641100).